https://dx.doi.org/10.22161/ijaers/3.12.34 ISSN: 2349-6495(P) | 2456-1908(O)

A Survey on the Project in title "Unsupervised Adaptive P300 BCI in the Framework of Chaotic Theory and Stochastic Theory"

Mohammed J. Alhaddad

Department of Information Technology, King Abdulaziz University, Faculty of Computing and Information Technology, Jeddah, Saudi Arabia.

Abstract—*In this paper we present a survey of work that has been done in the project "Unsupervised Adaptive P300 BCI in the framework of chaotic theory and stochastic theory"we summarised the following papers, (Mohammed J Alhaddad & 2011), (Mohammed J. Alhaddad & Kamel M, 2012), (Mohammed J Alhaddad, Kamel, & Al-Otaibi, 2013), (Mohammed J Alhaddad, Kamel, & Bakheet, 2013), (Mohammed J Alhaddad, Kamel, & Al-Otaibi, 2014), (Mohammed J Alhaddad, Kamel, & Bakheet, 2014), (Mohammed J Alhaddad, Kamel, & Kadah, 2014), (Mohammed J Alhaddad, Kamel, Makary, Hargas, & Kadah, 2014), (Mohammed J Alhaddad, Mohammed, Kamel, & Hagras, 2015).We developed a new pre-processing method for denoising P300-based brain-computer interface data that allows better performance with lower number of channels and blocks. The new denoising technique is based on a modified version of the spectral subtraction denoising and works on each temporal signal channel independently thus offering seamless integration with existing pre-processing and allowing low channel counts to be used. We also developed a novel approach for brain-computer interface data that requires no prior training. The proposed approach is based on interval type-2 fuzzy logic based classifier which is able to handle the users' uncertainties to produce better prediction accuracies than other competing classifiers such as BLDA or RFLDA. In addition, the generated type-2 fuzzy classifier is learnt from data via genetic algorithms to produce a small number of rules with a rule length of only one antecedent to maximize the transparency and interpretability for the normal clinician. We also employ a feature selection system based on an ensemble neural networks recursive feature selection which is able to find the effective time instances within the effective sensors in relation to given P300 event.*

The basic principle of this new class of techniques is that the trial with true activation signal within each block has to be different from the rest of the trials within that block. Hence, a measure that is sensitive to this dissimilarity can be used to make a decision based on a single block without any prior training. The new methods were verified using various experiments which were performed on standard data sets and using real-data sets obtained from real subjects experiments performed in the BCI lab in King Abdulaziz University. The results were compared to the classification results of the same data using previous methods. Enhanced performance in different experiments as quantitatively assessed using classification block accuracy as well as bit rate estimates was confirmed. It will be shown that the produced type-2 fuzzy logic based classifier will learn simple rules which are easy to understand explaining the events in question. In addition, the produced type-2 fuzzy logic classifier will be able to give better accuracies when compared to BLDA or RFLDA on various human subjects on the standard and real-world data sets.

Keywords— Type-2 fuzzy logic systems, linguistic model generation, modeling perceptions, brain computer interfaces.

I. INTRODUCTION

Mind reading is a human attempt to understand each other thoughts and feelings. It is very challenging task that relies on all our senses and fully exploits our cognitive and perceptual abilities. When we're trying to get inside someone's head, we comprehend the meaning of the words being spoken, we monitor facial expressions and body language, and we register the tone of voice and the cadence of speech. Until recently, the dream of being able to control one's environment through reading their minds had been in the realm of science fiction. However, today, humans can use the electrical signals from brain activity to interact with, influence, or change their environments.

The emerging field of Brain Computer Interface (BCI) technology may allow individuals unable to speak and/or use their limbs to once again communicate or operate assistive devices for walking and manipulating objects. BCI is an important tool that allows direct reading of information from the subject's brain activity by a computer. Such information can be used to perform actions controlled by the subject and hence provide an

https://dx.doi.org/10.22161/ijaers/3.12.34 ISSN: 2349-6495(P) | 2456-1908(O)

additional means of communication beside normal communication channels present in normal subjects. Such means can be the only way of communication with patients of such disease conditions as muscular dystrophy (MS) and therefore its development and enhancement have been the focus of many research groups in the past decade.

A BCI is a computer-based system that acquires brain signals, analyses them, and translates them into commands. Thus, BCIs do not use the brain's normal output pathways of peripheral nerves and muscles. The brain activity at different locations can be measured using different methods that include electroencephalography (EEG), magnetoencephalography (MEG), and some functional imaging modalities such as functional magnetic resonance imaging (fMRI). These techniques offer brain activity signal time courses that come from a particular location in the brain with the resolution of such spatial localization ranging from a few signals for the whole brain (as with EEG) to signal for each 1 mm3 voxel within the subject's brain (as with fMRI). The complexity of such systems also range from a simple, relatively inexpensive electrode cap worn by the subject and attached to a relatively small processing unit that provide very noisy signals while allowing subject mobility (as with EEG) to large expensive high field fMRI systems that allow excellent signal-to-noise ratio to be obtained while restricting the slightest subject motion during data acquisition. So, there is a clear trade-off between the quality of signals collected on one side and the mobility of the subject and the cost of the system on the other side. Approaches to improve quality of information from EEG-based systems through noise/artifact removal as well as more sophisticated analysis techniques would therefore allow this low cost, mobile technology to achieve better practical utility.

The user, often after a period of training, generates brain signals that encode intention. One of the important areas in BCI is to identify Event-Related Potentials (ERPs) which are spatial-temporal patterns of the brain activity which happen after presentation of a stimulus and before execution of a movement. One of the important ERPs is the P300 which is an endogenous component of ERPs with a latency of about 300 ms which is elicited by significant stimuli (visual, or auditory). There is a need to understand ERPs related phenomena like the P300 and their common characteristics across various humans and thus there is a need to develop easy to understand linguistic models explaining ERPs related phenomena like the P300.

Several machine learning based classification systems have been used in BCI. However, the vast majority of the employed techniques in BCI are black box models which are difficult to understand and analyse by a normal clinician. In addition, due to the inter and intra user uncertainties associated the P300 events, most of the existing classifiers need to be bespoke trained for the person using them and under given circumstances. However, if there is a change in the user or the given circumstances then the classifier need to be retrained.

Fuzzy Logic Systems (FLSs) have been credited with providing white box transparent models which can handle the uncertainty and imprecision. However, the vast majority of the FLSs employed in BCI were based on type-1 fuzzy logic systems which cannot fully handle or accommodate for the uncertainties associated with changing and dynamic phenomena such as the P300 in BCI whose amplitude and peak time change from one person to another. Type-1 fuzzy sets handles the uncertainties associated with the FLS inputs and outputs by using precise and crisp membership functions. Once the type-1 membership functions have been chosen, all the uncertainty disappears, because type-1 membership functions are totally precise. The uncertainties associated with BCI applications cause problems in determining the exact and precise antecedents and consequents membership functions during the FLS design. Moreover, the designed type-1 fuzzy sets can be sub-optimal for a given user and under specific age and health conditions. However due to the change in the individual circumstances and the uncertainties present between various people, the chosen type-1 fuzzy sets might not be appropriate anymore. This can cause degradation in the FLS classifier performance and we might end up wasting time in frequently redesigning or tuning the type-1 FLS so that it can deal with the various uncertainties faced. Type-2 FLSs which employ type-2 fuzzy sets can handle such high levels of uncertainties to give very good performances.

Several research articles addressed the problem of achieving higher quality of EEG signals for BCI applications with aim to improve the Signal-to-Noise Ratio (SNR).Two broad categories can be immediately recognized; namely, spatial domain techniques and temporal domain techniques. In the spatial domain techniques, the data from multiple spatially-distinct channels are utilized to identify the true signal projected onto all channels from the noise that is generally assumed to be independent among such channels. Such methods range from simple local spatial averaging to sophisticated variants of blind source separation methods such as independent component analysis, (Ramirez, Kopell, Butson, Hiner, & Baillet, 2011), (de Cheveigne & Simon, 2008), (Pires, Nunes, & Castelo-Branco, 2011), (Vorobyov & Cichocki, 2002), (Akhtar, Mitsuhashi, & James, 2012), (Geetha & Geethalakshmi, 2012). On the other hand, temporal domain techniques attempt to find similarities within the time domain of a single channel

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

signal that can be used to identify and suppress the noise components in that signal. This can be done by many methods ranging from simple averaging of consecutive epochs to transform domain based filtering techniques ranging from basic bandpass filtering (Mirghasemi, Shamsollahi, & Fazel-Rezai, 2006), (Hammon & De Sa, 2007) to different variants of the wavelet shrinkage method (Vázquez et al., 2012), (Ahmadi & Quiroga, 2013), (DL, 1995), (Quiroga & Garcia, 2003), (Effern et al., 2000), (Gao, Sultan, Hu, & Tung, 2010), (Saavedra & Bougrain, 2010), (Hammad, Corazzol, Kamavuako, & Jensen, 2012), (Sammaiah, Narsimha, Suresh, & Reddy, 2011), (Estrada, Nazeran, Sierra, Ebrahimi, & Setarehdan, 2011). A hybrid method between spatial and temporal methods has also been recently proposed to take advantage of available channels and redundant signal epochs (Tu et al., 2013). The predominant method of filtering used in BCI today is basic bandpass filtering that has become an essential part of the conventional pre-processing chain of BCI.

Even though previous denoising methods have contributed significant improvements, there are still limitations that need further research to reduce. For example, spatial domain methods rely on the availability of many channels (or electrodes), which would increase the cost, increase the weight, and cause loss of localization of EEG signals from the brain. Also, the integration of temporal domain signals into the pre-processing chain of BCI signals is yet to be done and is bound to increase the computational complexity requiring more expensive digital back-end hardware. Both techniques increase the power consumption of a portable BCI system due to additional channel front-ends or higher processing needed in the digital back-end. Therefore, a technique that would allow the use of a small set of channels and improve the performance of BCI system beyond the present methods at a reasonable computational cost would be highly desirable.

Brain-computer interface (BCI) offers hope for a communication channel with disabled patients who are not capable of using the normal communication channels. In spite of the major grounds covered by research in this area over the past decade, the challenge to make a reliable BCI system that combines mobility and accuracy remains open. Moreover, for some BCI techniques such as those based on detecting P300 signals in speller or one-of-multiple image selection tasks, existing commercial systems(intendiX) require assistance from caregivers or patient's family to operate the system by the patient. So, there is an immediate need for developing technologies that would lead to the availability of such devices to patients at home with more independence and in less restrictive settings.

In P300-based BCI, a signal is triggered by an auditory or visual stimulus when participants are asked to watch for a particular target stimulus presented within a stream of other stimuli in an oddball paradigm(Kübler & Müller, 2007). In all previous P300-based BCI interfaces, detection of the P300 is based on experience gained through calibration or training sessions prior to actual use to utilize supervised training sets to build the classification model (Hoffmann, Vesin, Ebrahimi, & Diserens, 2008). Several problems arise with this model including temporal variability of the signal (or inter-session variability) due to several reasons that include non-stationary brain dynamics and possible movement of electrode locations. This means that in practice, to communicate efficiently using such systems, the acquisition of data must be preceded by calibration with time difference as small as possible. As a result, the temporal persistence of such experience can be assumed to follow a training model close to an interpolation around the point at which the calibration/training was done, with most likely consequence is that the initial accuracy is expected to fade as time goes by after the initial training session. This required training imposes limitations on the utility (and hence commercialization) of the technology by individuals outside of research labs. Hence, efforts must be directed to develop methods that use unconventional decision models to overcome such limitations and achieve sufficient robustness for practical utility.

Previous work attempted to decrease the amount of calibration required for BCI and move toward a zero-training goal (Krauledat, Tangermann, Blankertz, & Müller, 2008). This method relies on observing the variation in training sessions and fitting such variation to spatial filters that can be used to make calibration sessions shorter. Even though the goal of this method is zero training, the approach relies on the utilization of prior information to allow future calibrations to be shorter or ideally no required. In that sense, it can be considered as a supervised training method with a more efficient training strategy that makes it possible for the training model to be more generalized and thus last longer. As a result, developing a P300-based BCI technology that can work adaptively without any prior calibration is still an open goal that once achieved would further the use of this important technology in real-life.

The project has two main contributions as follows:

Develop a denoising method for P300-based brain-computer interface data that allows better performance to be obtained with lower number of channels and blocks. The new method will be applied to experimental data and compared to the classification results of the same data using the same pre-processing and classification steps to allow direct comparison of results. Also, the new method

https://dx.doi.org/10.22161/ijaers/3.12.34 ISSN: 2349-6495(P) | 2456-1908(O)

will be compared to bandpass filtering and wavelet shrinkage based denoising as the relevant and widely used method for denoising at the present. Performance in different experiments will be quantitatively assessed using classification block accuracy as well as bit rate. The computational complexity of the new method is also described and compared to previous methods.

Investigating new methodologies for P300-based brain-computer interface data that require no prior training. This targets the development toward "plug-and-play" P300-based BCI devices whereby the device is taken out of the box and used immediately by the patient. The new method will be applied to experimental data and compared to the classification results of the same data using the conventional processing techniques requiring prior lengthy training sessions.

Developing a genetic type-2 fuzzy logic based classifier which is able to handle the inter and intra user uncertainties to produce better prediction accuracies when compared to other competing classifiers such as Bayesian Linear Discriminant Analysis (BLDA) and Regularized Fisher Linear Discriminant Analysis (RFLDA). In addition, the generated type-2 classifier is learnt from data via a genetic algorithm to produce a small number of rules with a rule length of only one antecedent to maximize the transparency and interpretability for the normal clinician. We also employ a feature selection system based on an ensemble neural networks recursive feature selection which is able to find the effective time instances within the effective sensors in relation to given P300 event. We will present various experiments which were performed on standard data sets and using real-data sets obtained from real subjects experiments performed in the BCI lab in King Abdulaziz University. It will be shown that the produced type-2 fuzzy logic based classifier will learn simple rules which are easy to understand explaining the events in question. In addition, the produced type-2 fuzzy logic classifier will be able to give better accuracies when compared to BLDA or RFLDA on various human subjects on the standard and real-world data sets.

II. METHODOLOGY

2.1 Spectral Subtraction Denoising

The methodological approach that will be followed in this work is to adopt spectral subtraction based signal denoising, which is an effective speech signal denoising method that was previously applied to fMRI signal denoising(Kadah, 2004). This method uses adaptive estimation of noise and does not assume a model for the true signal thus matching well our problem. Here, we derive the spectral subtraction method for EEG applications and point out the modifications to the previous work to meet our unique application requirements.

Using the traditional additive noise model, the EEG temporal signal can be modelled as the summation of a true response signal, a physiological/instrumentation baseline fluctuation component, and a random noise component (Kadah, 2004). The physiological/instrumentation baseline fluctuation component can be considered as a deterministic yet unknown signal such as baseline drift or physiological motion artifacts and can be dealt with using existing pre-processing methods(Hoffmann et al., 2008). On the other hand, the random part consists of two components: the thermal noise in the electronics of the data acquisition system and the superimposed signals from neighbouring neurons not involved in the true response sought. While the former component is well known to be Gaussian white noise process, the latter can also be shown to be so using a straightforward application of the central limit theorem to the summation of many signals of random activation patterns. Therefore, we will assume an additive noise model whereby the measured signal is practically the sum of a deterministic component d(t), including both the true EEG signal and low frequency or baseline wander, in addition to an independent random noise n(t). That is,

$$s(t) = d(t) + n(t). \quad (1)$$

Given that d(t) and n(t) are independent, the power spectrum of the measured signal can be given as,

$$P_{ss}(\omega) = P_{dd}(\omega) + P_{nn}(\omega). \quad (2)$$

Hence, the power spectrum of the deterministic part of the signal can be theoretically computed as (Kadah, 2004),

$$P_{dd}(\omega) = P_{ss}(\omega) - P_{nn}(\omega). \quad (3)$$

So, the deterministic signal power spectrum is obtained by subtracting the spectra of the measured signal and an estimate of the random noise power spectrum. Practically speaking, to estimate that deterministic signal itself from the above estimated power spectrum, the magnitude of its frequency domain can be directly computed as the square root of the power spectrum. However, we need to find the phase part as well in order to be able to inverse-transform the frequency-domain estimate back to the time-domain signal. Several techniques can be used to do that. One such method relies on an estimate obtained from the phase of the Fourier transform of the original signal S(ω). In this case, the spectrum of the estimated deterministic signal Sd(.) can be given as (Kadah, 2004),

$$S_d(\omega) = \sqrt{P_{dd}(\omega)} \cdot \exp(j\,\mathrm{Phase}(S(\omega))) \quad . \quad (4)$$

The denoised deterministic signal sd(t) is then computed as the real part of the inverse Fourier transformation of this expression. A block diagram of this method is shown in Figure 1

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

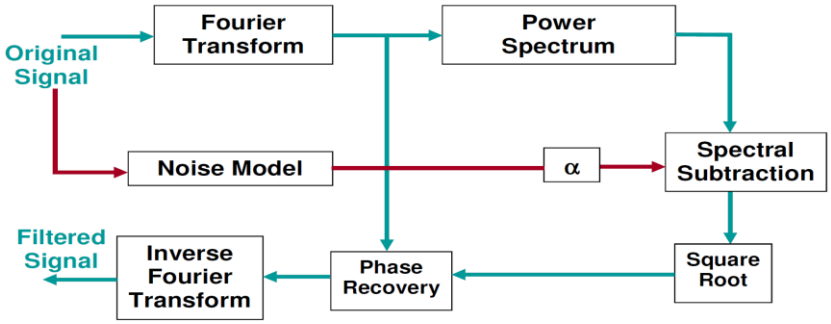


Fig.1: Original Spectral Subtraction denoising block diagram. The same data is used to estimate the noise power spectrum which is then removed from the overall power spectrum

In spite of the success of this method in denoising event related functional magnetic resonance imaging time courses, two problematic issues are present in our application to EEG recordings. The first is the use of the phase component of the original signal in the denoised signal. Given that the information in the phase part is as important as that in the magnitude part, leaving this component intact will no doubt limit the efficiency of the process in removing the noise components in the final signal. This issue was also a concern in the original application of this technique in fMRI and it was found that the improvement is still robust and therefore this issue is not as critical. The second issue is related to the observed jumps between the initial and final time points in the EEG epochs due to such effects as baseline drift that was found to be present and in many cases severe in the data sets we used in this work and elsewhere. This is an important difference between our case and the application of this method to fMRI signals where baseline wander is present but much less severe. Such large differences between first and last points in EEG epochs introduce incorrect high frequency components in the estimated power spectrum as a direct result of the Discrete Fourier Transform (DFT) model. The DFT assumes the measured epoch to be one period of a periodic signal, which means that the transform will see sharp discontinuities at both borders of the signal. As a result, this causes artifacts in the denoised signals that are deterministic yet unknown depending on the magnitude of such variable jump. This makes this technique not acceptable as a valid pre-processing tool in this application because of its introduction to such systematic errors. An illustration of such artifact is given in Figure2 where the first part of a sample EEG epoch is shown before and after the old spectral subtraction processing. It can be observed that the beginning part of the denoisedsignal shows a clear artifact.

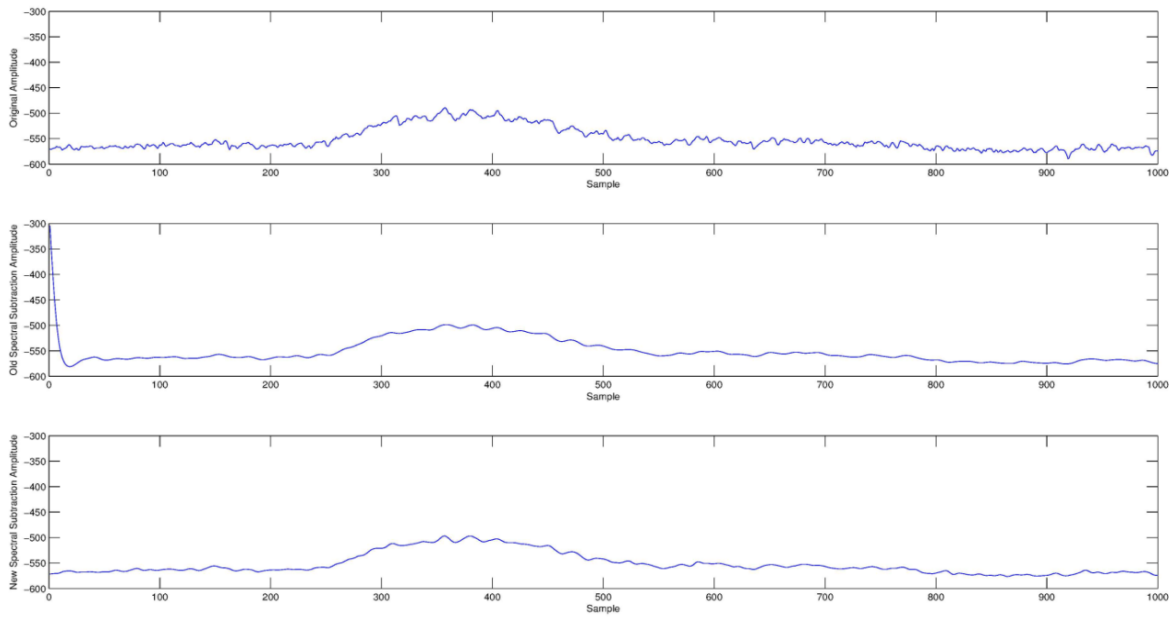


Fig.2: Illustration of border discontinuity artifact in the old spectral subtraction method and its solution in the new method. The top plot shows the original signal, the middle plot shows the signal processed with the old spectral subtraction method with a clear artifact at the border in the beginning of the signal. This problem is absent in the new spectral subtraction method at the bottom.

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

To solve the above problem and allow artifact-free use of spectral subtraction, we here propose a modified version of the spectral subtraction method in which the original signal is converted to an even-symmetric signal by concatenating the signal with its time reversed version before using the discrete Fourier transform to estimate the power spectrum. This bears similarity to what is done in the widely-used discrete Cosine transform. This has two important implications that address the above two issues in the original method. First, the phase of this even-symmetric signal is expected to be zero for positive frequency amplitudes or π for negative ones. However, we observe a deterministic linear phase corresponding to a shift of ½ point since the origin of symmetry of this signal lies in between the two middle points. This changes the role of the phase estimation in the original method to merely sign detection and compensation for the deterministic ½ point shift yielding very high noise immunity. Second, the even symmetric signal form ensures the continuity at both ends of the signal to be preserved thus eliminating edge artifacts. The block diagram of the modified version of spectral subtraction is presented in Figure 3. The result of the using the modified spectral subtraction on the same signal in Figure 2 is shown at the bottom plot where the artifact present in the old spectral subtraction method is completely absent in the new method.

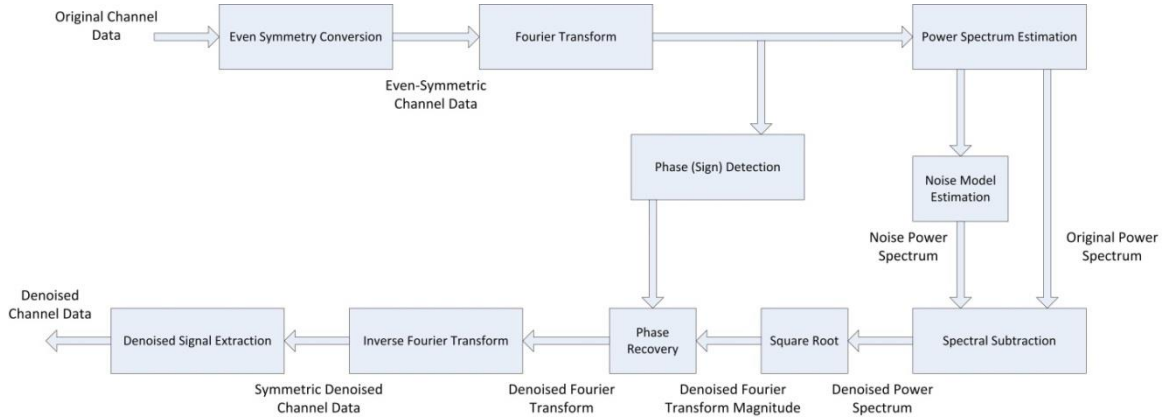


Fig.3: Block diagram of Modified Spectral Subtraction Denoising where the data are concatenated with its mirror image to generate a symmetric signal before applying the regular steps of the spectral subtraction method. This allows the phase of the signal to be zero and avoids artifacts from mismatch of signal levels at the borders.

The detailed steps of implementation of the new method are given as follows:

Step 1: Read in the raw epoch data s(t) and convert it to a symmetric signal by concatenation with its reflected version s(-t).

Step 2: Compute the fast Fourier transform of the symmetric raw epoch data. Estimate and keep the linear phase of the result.

Step 3: Compute the periodogram-based estimate of the power spectrum as the squared magnitude of the fast Fourier transform of the raw epoch data.

Step 4: Estimate the noise level by computing the average of the power spectrum values in the upper 20% of the frequency range that contains no signal components.

Step 5: Use Equation (3) to compute the power spectrum of the denoised signal. If the subtraction result at any frequency is negative, it is clipped to zero.

Step 6: Compute the denoised signal discrete Fourier transform as the square root of the denoised signal power spectrum and transform it back to the time-domain denoised signal after adding the deterministic linear phase estimated in Step 2.

In order to implement the above denoising strategy, the noise power spectrum has to be estimated. Given that the noise model is Gaussian white noise, its power spectrum is well known to be constant over all frequencies that is directly proportional to the noise variance. Hence, it is sufficient to estimate a single parameter in order to completely determine the noise power spectrum.

Our strategy in this work is to have the new denoising technique implemented as a transparent block that can be used with existing trial extraction and pre-processing methods without any modifications to the other blocks. Therefore, we insert the new denoising block in between reading the session data file and the referencing step where the individual channel signals are read and processed using the new method then passed on to further processing steps in the same format they were read (as shown in Figure 4). Since this method should work adaptively, the estimation of the noise variance must be done adaptively from the original signals without any user intervention.

https://dx.doi.org/10.22161/ijaers/3.12.34 ISSN: 2349-6495(P) | 2456-1908(O)

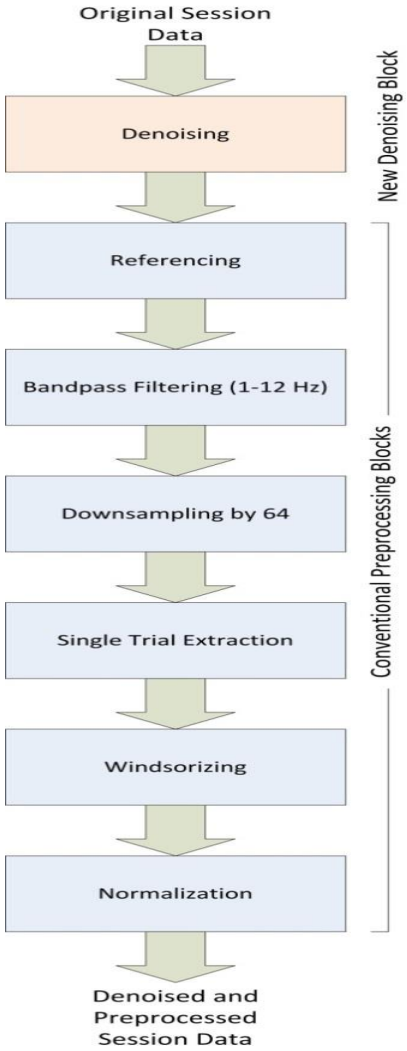


Fig.4: Block diagram of proposed new pre-processing chain with an added new denoising block before the usual steps conventionally applied to BCI signals.

This was done as follows. Since the original channel data are recorded using a much higher sampling rate than needed for the known frequency content of EEG signals and what is conventionally used for activation detection, the power spectrum of the original signal can be assumed to have noise only in its high frequency components. Consequently, the noise level can be estimated directly from the power spectrum of the original signal as the average of the upper half of the power spectrum as shown in Figure 5.

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

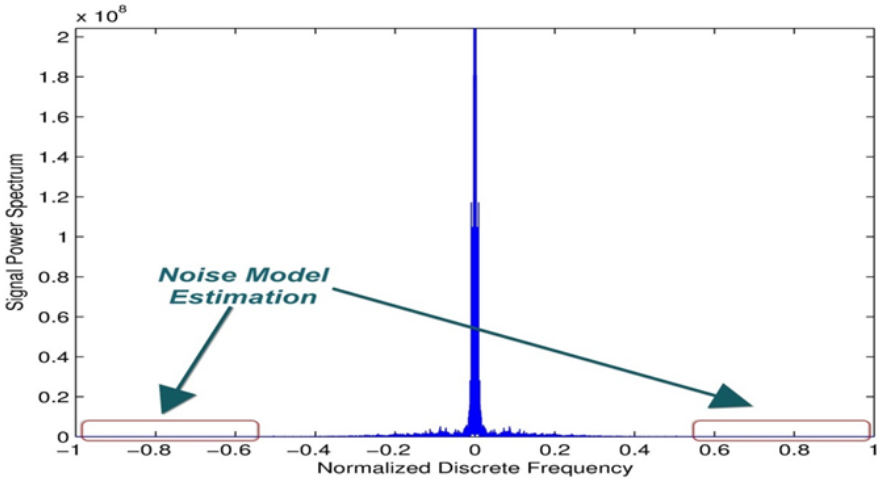


Fig.5: Illustration of noise power spectrum estimation from the upper part of the signal power spectrum on both ends known to have no true signal components based on the average of such areas.

The average is used because the power spectrum itself at each point can be shown to be a random variable that is unbiased (that is, mean is equal to true value) and consistent (variance decreases uniformly to zero as number of points goes to infinity). Given that the magnitudes of these points are independent and identically distributed, their average can be used to improve the estimation of the common mean of their processes. This estimation process is done for each channel independently and used to denoise its respective channel to account for different analog front-ends for each channel.

2.2 Unsupervised Processing Methods

The basic principle of the new class of unsupervised techniques is that the trial with P300 signal within each block has to be different from the rest of the trials within that block. In other words, if we have an N-trial block, one signal has a different form from all the other (N-1) signals. This means that if we could find a measure that can be sensitive to this dissimilarity (or alternatively, the similarity of signals without activation), then we can indeed make a decision based on a single block without any prior training. In the following, we present a number of such measures and discuss how they will be used to separate the activated trial from the rest of trials in the same block. The input to each of these methods is a set of trials $\{x_1, x_2, \dots, x_N\}$ given as a collection of M×1 vectors. The general block diagram for all methods is presented in Figure 6.

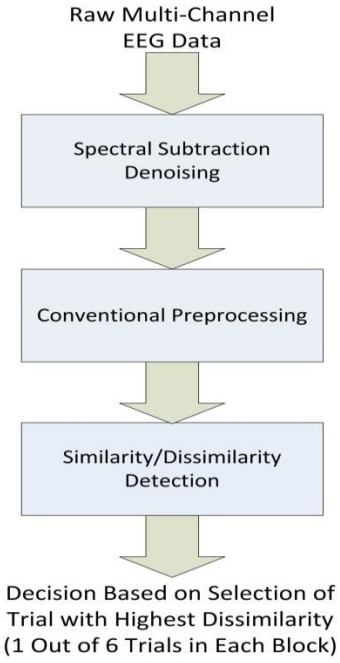


Fig.6: Block diagram of the new unsupervised approach.

Outlier Detection Method

In this method, each trial is formulated as a vector in M-dimensional space where M is the number of points within each trial. The assumption underlying this method is that the vectors of all trials that exhibit no P300 activation will be similar and that they are all different from the one that has a P300 activation. Hence, a distance measure is used to compute the distance between each trial and all other trials in a pairwise manner. Then, for each trial, the sum of all distances with other trials is used to differentiate the one trial with the largest distance from all other trials. In mathematical form, the trial with P300 signal is computed as the solution to the optimization problem given by,

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

$$\max_i \sum_{\text{all } j \neq i} \|x_i - x_j\|. \quad (5)$$

This means that this method detects the trial that is the furthest from all other trials. The norm in this equation was used as the 2-norm (Strang & Press, 2009). Possibility of using other norm definitions is possible but this one was selected to make its concept more visible by appealing to the common Euclidean distance as the measure used in this problem.

Correlation Method

The detection of the P300 signal relies on its characteristic shape and onset that are unique and help distinguish this type of activation from any other type. It is also very common for the literature working with P300 signals to show the P300 signal form their data by simple averaging of a number of trials with known activation presence. Here, we borrow an activation detection method from functional magnetic resonance imaging given the similarity between this technique and that of P300 based BCI. In particular, if the activation signal shape is somewhat known, it is possible to detect its presence by simple correlation of a "template" activation and each trial signal. If we have N trials within a block, then the trial with the strongest correlation with the template activation should most likely be the one with P300 signal. Since the onset of the true P300 signal varies between 300 and 500 ms, the template activation is used with different amounts of time shift to detect such correlation to make sure that such variability is taken into consideration. In a mathematical form, the activated trial is found by solving the following optimization over all trials in the block:

$$\max_{\text{all } i,\Delta t} \{x_i^T s_{\Delta t}\}. \quad (6)$$

Here, $s_{\Delta t}$ is the template activation signal shifted in time byΔt. It is possible to constrain the range of time shifts to include only those with onset within the known range of P300 signal. However, this was not done in this work and the range of shifts was extended to be the full range of [-M, M] for the M-sample signals.

Dot Product Method

This method is very similar to the Outlier Detection method above with the only exception in that the measure is here the dot product of the two trial vectors rather than the norm of their difference. This dot product relies on the fact that the dot product designates the component of one vector onto the other or basically the cosine of their angle if they both have similar magnitudes (Golub & Van Loan, 2012). That is, similar vectors have higher dot products and vice versa. So, the optimization is here to find a trial that has the smallest dot product with all remaining vectors. Consequently, Eq. (1) is modified to be as follows:

$$\min_i \sum_{\text{all } j \neq i} x_i x_j. \quad (7)$$

Cross-Correlation Method

This method bears similarities to both the dot product method and the correlation method. In particular, rather than computing the dot product between two trials, it computes the cross-correlation between them and obtains the peaks of this cross-correlation function as the measure of similarity for this method. So, it is a generalization of the concept in the dot product method and also a variant of the correlation method whereby the template signal is just a different trial in the same block. In a mathematical form, we select the trial that satisfies the following optimization problem:

$$\max_i \left\{ \sum_{j \neq i} \max_{\Delta t} x_i^T x_{j,\Delta t} \right\}. \quad (8)$$

Singular Value Decomposition (SVD) Method

The issue of representation of a set of vectors is a well-known problem in mathematics and also has wide utility in many applications. Some of the well-known solutions are based on the Principal Component Analysis (PCA) that allows the computation of the so-called "principal component" that best represents a set of vectors by inspecting the eigenvalues of the different eigenvectors in the Eigen-decomposition of the problem and finding the eigenvector with much larger eigenvalue from the rest. As the set of vectors become more and more independent, it becomes more difficult to find a single vector that can best represent them all ending with the ideal case of orthonormal basis that result in all unity eigenvalues. In our context, the assumption is that (N-1) trials are somewhat similar (at least not independent). Therefore, if we perform such analysis for each of the possible (N-1) trials and using a sparsity measure for the resultant eigenvalues of each decomposition that can detect how close each of these sets of vectors to the idea case of a single outstanding eigenvalue, it can be possible to detect the trial with activation as the remaining vector (Hyviirinen, Karhunen, & Oja, 2001). That is, the most sparse set of eigenvalues of all decomposition denote that these trials are not activated and it turn the remaining vector has the P300 signal. The 1-norm measure was selected as the sparsity measure for the singular values in our implementation. In a mathematical form (Golub & Van Loan, 2012):

$$\text{For Trial i}, A_i = \{x_j\}_{1 \leq j \leq N, j \neq i} = U \Sigma V^T, \quad (9)$$

where U and V are orthogonal matrices of size MxM and (N-1)x(N-1) respectively, and Σ is a Mx(N-1) matrix with its upper (N-1)x(N-1) matrix taking a diagonal form with singular values sji on the diagonal and the lower (M-N+1)x(N-1) a zero matrix. The activated trial is hence

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

taken as the solution to the following optimization problem,

$$\max_{i\ ,\text{all } A_i} \|\{s_1, s_2, \dots, s_{N-1}\}\|_1 \quad (10)$$

That is, we select the trial that has all the remaining trials forming a matrix with the largest 1-norm for its singular values (similar to the strategy used with compressive sensing (Candè & Wakin, 2008).

2.3 A Genetic Interval Type-2 Fuzzy Logic Based Approach for Generating Interpretable Linguistic Models for the Brain P300 Phenomena

Several machine learning based classification systems have been used in BCI where linear and non-linear classification methods have been employed including Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) (Lotte, Congedo, Lécuyer, Lamarche, & Arnaldi, 2007), (Müller, Anderson, & Birch, 2003), (Croux, Filzmoser, & Joossens, 2008). Popular LDA include Fisher's Linear Discriminant Analysis (FLDA) and the regularized versions termed Regularized Fisher Linear Discriminant (RFLD) as well as Bayesian Linear Discriminant Analysis (BLDA). The regularized version of FLDA may give better results for BCI than the non-regularized version Lotte, 2007 #35}, (Müller et al., 2003), (Croux et al., 2008). However, these techniques are black box models which are difficult to understand and analyze by a normal clinician. In addition, most of these classifiers need to be bespoke trained for the person using them and under given circumstances. However, if there is a change in the user or the given circumstances then the classifier need to be retrained.

Fuzzy Logic Systems (FLSs) have been credited with providing white box transparent models which can handle the uncertainty and imprecision. FLSs have been used in (Lotte, Lécuyer, & Lamarche, 2007) for motor imagery classification in BCIs which produced similar results to the most popular classifiers used in BCIs while providing an easy to read and interpret model. The work on (Lotte, Lécuyer, & Arnaldi, 2009) employed FuRIA which is a fuzzy logic based trainable feature extraction algorithm for BCIs which is based on inverse solutions. This algorithm can be trained to automatically identify relevant regions of interest and their associated frequency bands for the discrimination of mental states. The main drawback of FuRIA is its long training process (Lotte et al., 2009). Palaniappan et al. (Palaniappan, Paramesran, Nishida, & Saiwaki, 2002), presented a new BCI design using fuzzy ARTMAP neural network whose objective was to classify the best three of the five available mental tasks for each subject using power spectral density of EEG signals. The suggested fuzzy based system has been used successfully with a tri-state switching device. However, all these FLSs were based on type-1 fuzzy logic systems which cannot fully handle or accommodate for the uncertainties associated with changing and dynamic phenomena such as the P300 in BCI whose amplitude and peak time change from one person to another. Type-2 FLSs which employ type-2 fuzzy sets can handle such high levels of uncertainties to give very good performances.

Recently type-2 FLSs have been applied in BCI applications where Herman et al. (Herman, Prasad, & McGinnity, 2008) presented an Interval Type-2 FLS (IT2FLS) classifier design methodology so that the BCI system non-stationaries can be effectively handled. In(Herman et al., 2008), an initial rule base structure was first initialized and then the system parameters were globally optimized. The IT2FLS has been applied to the problem of classification of Motor Imagery (MI) related patterns in EEG recordings which was regarded as a major difficulty for the state-of-the-art BCI methods (Herman et al., 2008). It was found that the IT2FLS resulted in better performance when compared to T1FLS, LDA and SVM which are commonly utilized in BCI systems.

The Employed Feature Selection Techni

After signal pre-processing, we need to perform feature selection to identify the most important features contributing to the BCI output. In this paper, we will use a different feature selection method from what is used in the BCI literature. The traditional feature selection techniques like Principle Component Analysis (PCA) are used to reduce complex data with a large number of attributes into lower dimensions to determine subtle features within the data. These approaches however do not provide a means of showing the degree of influence and affect each input feature has on the output.

We will use a neural networks based approach for feature weighting which runs a recursive feature elimination scheme. We have chosen this neural networks based feature selection as we are interested to know the relevance of each input and its weight in relation to the BCI output, thus having a justification for the feature selection decision. Hence, the employed neural networks based techniques, unlike other feature selection methods, not only extract the important and relevant input features, but the employed neural networks based method can also identify the degree of influence and effect each input feature has on the output (i.e. the weight of the given important input features).

Neural Networks are able to learn and adapt from training noisy data and they are capable of acting as universal approximates. Once trained, they provide fast mapping from inputs to outputs. Neural networks therefore have the potential to better capture the most relevant features related to a classification task.

We follow the feature selection mechanism presented in (Windeatt, Duangsoithong, & Smith, 2011) which

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

operates recursively in two steps. First rank the features according to a suitable feature-ranking method and then identify and remove the r least ranked features. As in (Windeatt et al., 2011), we used feature ranking by ensemble neural network MultiLayer Perceptron (MLP) weights combined with recursive feature elimination. The output O of a single output single hidden-layer MLP, assuming sigmoid activation function S is given by

$$O=\sum_q S\sum_P(x_p W_{pq}^1) * W_q^2 \quad (11)$$

Where p,q are the input and hidden node indices, xp is input feature, W1 is the first layer weight matrix and W2 is the output weight vector. After the MLP training, the feature weighting is extracted from the trained network is as follows: for an input node m its feature weight is given by(Windeatt et al., 2011):

$$wp=\sum_q \left|W_{pq}^1 * W_q^2\right| \quad (12)$$

The weight wp in Equation (12) is the sum over hidden nodes of the product of two weights connected via each hidden node to the pth feature. As we are using Ensemble MLPs of the 10 neural networks, then each feature will have a weight wp for each MLP. Then individual rankings are averaged and scaled for each feature, giving an overall % ranking, which is used for eliminating the set of least relevant features at each recursive step. We eliminate at each recursive step 10 least ranked features. The process continues where the accuracy of prediction is within certain a range as this means cutting redundant or noisy features is not affecting the accuracy of prediction much. However, when the accuracy of prediction drops drastically, the feature selection process is stopped as the accuracy drops significantly when important features are cut down.

Figure 7 shows the list of the extracted features and their corresponding weights for the P300 BCI application which will be shown in the experiments section. For this application, the pre-processed data from four electrodes sensors were fed to the feature selection. The pre-processed data consists of 32 readings per each second for each electrode sensor.

Rank	Feature Description	Weight
1	Third Sensor, 6th Time instance	91.3%
2	Fourth Sensor, 26th Time instance	89.2%
3	Second Sensor First Time Instance	84.1%
4	Fourth Sensor, 4th Time instance	83.7%
5	Second Sensor 12th Time Instance	83.3%
6	Second Sensor 5th Time Instance	81.0%
7	Fourth Sensor 29th Time Instance	80.9%
8	Second Sensor 12th Time Instance	79.8%
9	Second Sensor 8th Time Instance	79.7%
10	Third Sensor, 14th Time instance	78.5%
11	Third Sensor 7th Time Instance	78.5%
12	Third Sensor 19th Time Instance	78.0%
13	Third Sensor 24th Time Instance	77.1%
14	Third Sensor 21st Time Instance	76.9%
15	Fourth Sensor 12th Time Instance	76.4%
16	Third Sensor 17th Time Instance	76.2%
17	First Sensor 16th Time Instance	75.8%
18	First Sensor 18th Time Instance	75.7%
19	First Sensor 8th Time Instance	75.0%
20	Third Sensor 4th Time Instance	74.2%
21	First Sensor 23rd Time Instance	73.8%
22	Fitst Sensor 7th Time Instance	73.6%
23	Fouth Sensor 28th Time Instance	72.9%
24	First Sensor 6th Time Instance	70.0%
25	Fourth Sensor 18th Time Instance	69.7%
26	First Sensor 16th Time Instance	69.2%
27	Third Sensor 28th Time Instance	68.1%
28	First Sensor Second Time Instabce	68.1%
29	Thitd Sensor 4th Time Instance	67.3%
30	Second Sensor 12th Time Instance	67.3%
31	Fouth Sensor 14th Time Instance	67.1%
32	Fourth Sensor 29th Time Instance	66.9%
33	Fourth Sensor 22nd Time Instance	66.6%
34	First Sensor 29th Time Instance	66.4%
35	First Sensor 29th Time Instance	66.4%
36	First Sensor 31st Time Instance	65.9%
37	Second Sensor 13th Time Instance	65.9%
38	Second Sensor 22nd Time Instance	63.6%

Fig.7: The features extracted for a P300 BCI application with 4 electrodes sensors.

Figure 7 shows that only 38 features were selected as they were found to be the least number of features to give good prediction accuracy. Figure 7 enables us to have a transparent interpretation of what happens in the brain where it shows that the P300 related phenomena was mainly affected by the Fourth, Third and Second Sensor and that the First Sensor started appearing from the 17th feature onwards which implies that it is not as important

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

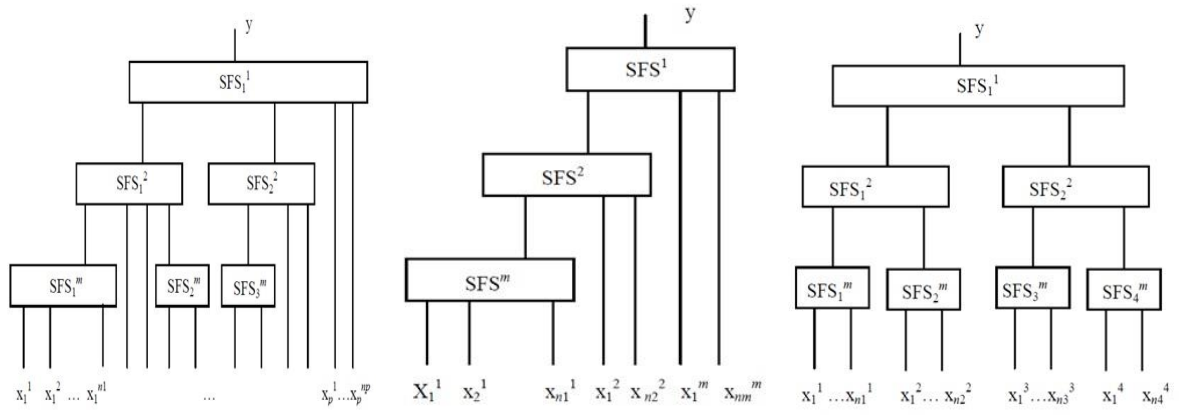
as the other three sensors. The List in Figure 7 also allows us to see which time instances per second are crucial for each electrode sensor

The Proposed Genetic Interval Type-2 Hierarchical Fuzzy Logic Classifier for P300 BCI Application

In this section, we will present a novel genetic hierarchical type-2 fuzzy logic classifier for BCIs. The proposed classifier will aim to:

- Generate a white box model which has a relatively small set of rules where the length of each rule is kept to a minimum of one antecedent. This will allow having linguistic models which are easy to understand and analyse by the normal clinician. This model will also enable us to have more insight into the learning operations happening in the brain.
- Handling the fuzzy logic known curse of dimensionality problem by implementing the system in a hierarchical way which helped in reducing the number of rules. This made it easier to optimize the fuzzy logic based system without affecting the system's ability to handle inputs' uncertainties.
- Employ type-2 fuzzy logic systems to handle the encountered uncertainties within the P300 phenomena where there exist high levels of inter and intra users uncertainties which could be due to psychophysical state of the subject, physical layout of the stimuli, sequence of stimuli and the variability of the P300 time and amplitude from one person to another, etc. The type-1 fuzzy logic models will not be able to handle and model such uncertainties. Hence, the interval type-2 fuzzy logic sets through their FOU will be able to handle and model the encountered uncertainties to result in a novel model enabling to model the complex uncertainties taking place within the human which will allow to generate a general model explaining the nature of a brain phenomenon rather than specific patterns bespoke for certain users in certain circumstances.

To satisfy the first and second objective of generating a small number of rules, we will employ the hierarchical structure shown in Figure 8.



(a) **(b)** **(c)**

Fig.8: The General Hierarchical Fuzzy Logic Structure. ***(b)*** *Incremental hierarchical Structure.* ***(c)*** *Aggregated hierarchical Structure .*

To illustrate how the interval type-2 hierarchical fuzzy logic classifier reduce the number of rules compared to other fuzzy logic classifiers, we will give the following example: Let's assume that the classifier extracted the maximum number of possible rules from the data where each rule is represented by one antecedent and one consequent and all the inputs are numerical and are represented by the same number of fuzzy sets. In order to calculate the maximum number of possible rules, we will use the following equation:

$$R = N \times S \quad (13)$$

Where:

- R represents the number of rules
- N represents the number of inputs
- S represents the number of fuzzy sets in each input

In a normal fuzzy logic system, the number of rules is given as follows:

$$R = S^N \quad (14)$$

The hierarchical fuzzy logic systems shown in Figure 8 address the "curse of dimensionality" problem but they face the intermediate inputs and rules problem where it is not easy to define rules for them. On the other hand, the proposed hierarchical fuzzy logic system shown in Figure 9 addresses the "curse of dimensionality" problem where its rules grow linearly with the number of inputs and it

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

has no intermediate inputs and rules. In addition, the proposed hierarchical fuzzy system can be easily designed and it is modular where low levels FLSs could be easily added or deleted.

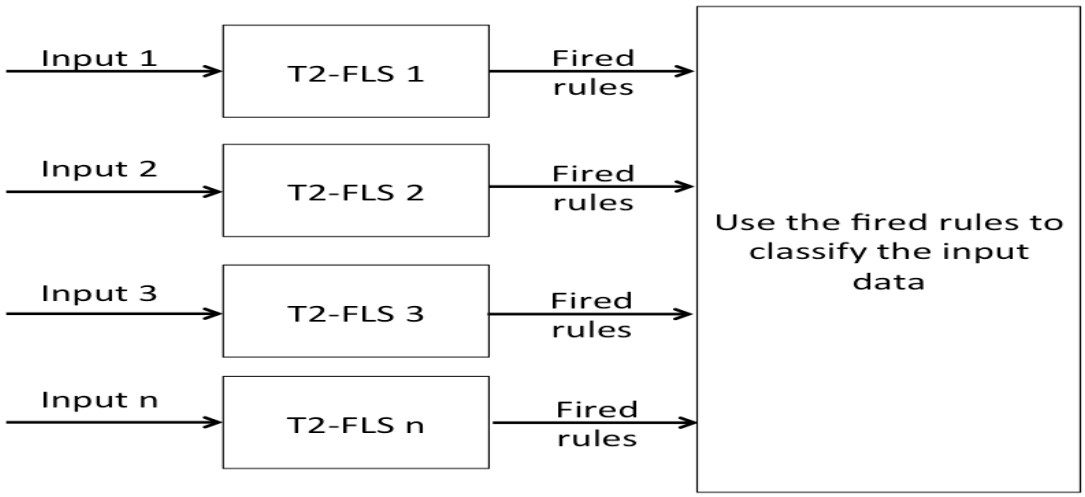


Fig.9: The proposed Interval Type-2 Hierarchical Fuzzy Logic Classifier structure

Genetic Learning of the Rules of the Type-2 Based Classifier

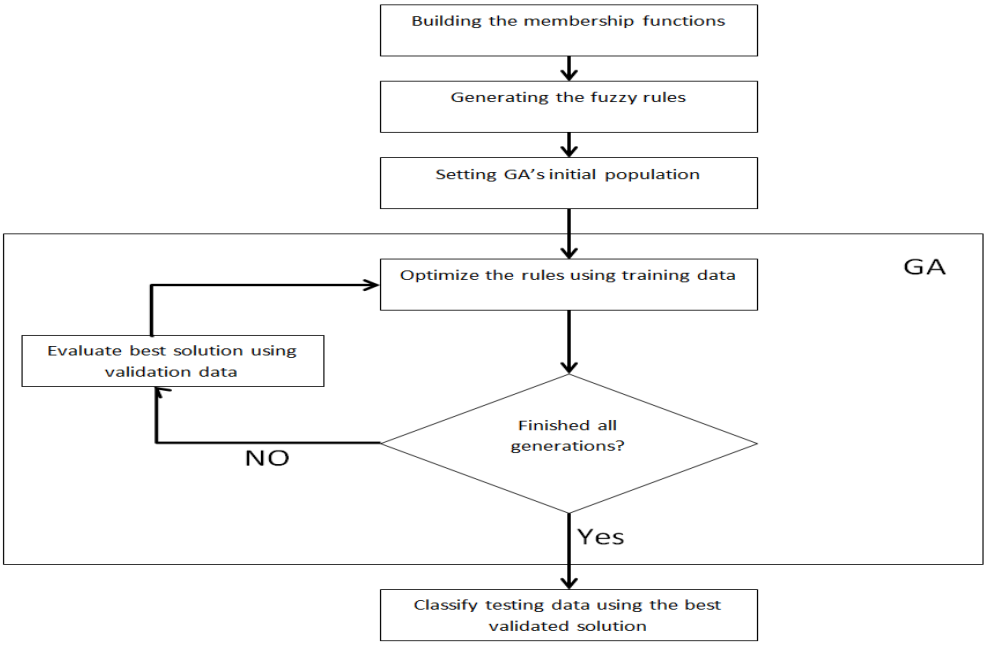


Fig.10: An overview of the proposed genetic interval type-2 hierarchical fuzzy logic based classifier.

Figure 10 shows an overview on the proposed genetic type-2 hierarchical fuzzy logic based classifier. The proposed classifier involve the following stages:

- Building equally spaced type-2 membership functions for each input using the input's minimum and maximum values.
- Generating the fuzzy rules using the training data
- Setting the Genetic Algorithm (GA) initial population
- Optimize the fuzzy rules using GA and evaluate it using the training data. Evaluate the best GA solution using validation data and store the best solution. Classify the testing data using the best validated solution.

The classifier involves the setup of the following parameters:

- Number of GA generations (we have employed 500 generations)
- The size of the population in each generation (the population was made of 100 chromosomes)
- Crossover probability (we have employed a crossover probability of 0.8)
- Mutation probability (we have employed a mutation probability of 0.1)
- The number of fuzzy sets that will be used in defining the inputs (we have used two fuzzy sets per input)

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

- The percentage that will be used in dividing the data into training, validation and testing (the division employed in this work is 70 % of data for training (of which 10% are used for validation) and 30 % for testing).
- It should be noted that all the GA parameters were chosen through empirical experiments in order to allow the fuzzy classifiers to achieve maximum accuracy. As will be shown later the GA optimises the rule base while the type-2 fuzzy sets will be obtained as explained below. In the following subsections, we will describe in detail each of the above steps.

Building Equally Spaced Type-2 Membership Functions for Each Input Using the Input's Minimum and Maximum values

For each input, the minimum and maximum values of this input are used to build equally spaced membership functions for this input.

Let $\tilde{A}_f^n$ be an interval type-2 fuzzy set where $n = 1 \ldots N$, N is the number of inputs and $f = 1 \ldots F$, F is the number of fuzzy sets. We will distinguish between two cases where $F > 2$ and $F = 2$ as follows:

Case 1 ($F > 2$): the universe of discourse for input x_n is determined by min (x_n) and max (x_n) as can be seen in Figure 11. We calculate $step_n$ to divide the universe of discourse into equally spaced intervals as follows:

$$step_n = \frac{\max(x_n) - \min(x_n)}{F - 1} \qquad (15)$$

As illustrated in Figure 11, for f=2,…F-1 we calculate $start_f^n$ as follows:

$$start_f^n = \frac{step_n}{2} + (step_n \times (f - 1)) \qquad (16)$$

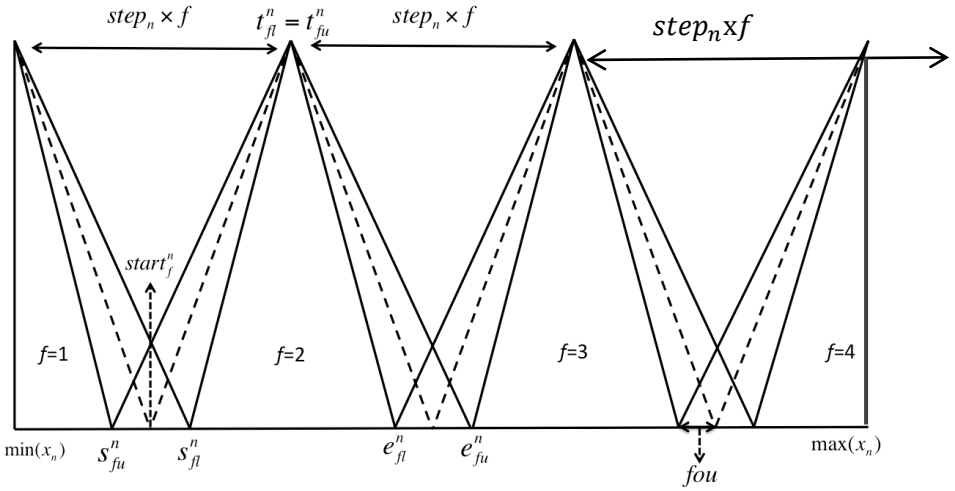


Fig.11: Building Equally Spaced membership function.

The core of the type-2 fuzzy sets are the same as the type-1 fuzzy sets, the FOU of the interval type-2 fuzzy sets is formed by blurring the left hand and right end points of the base of the type-1 fuzzy sets by a value called dfou which represents the blurring factor to generate the FOU of interval type-2 fuzzy sets.

$$dfou = {(max(x_n) - min(x_n))}/{(4 \times (F - 2))} \qquad (17)$$

The Interval type-2 fuzzy set $\tilde{A}_f^n$ is bounded by a lower bound type-1 membership function and an upper bound type-1 membership function. The upper membership function is represented by three points $[s_{fu}^n, t_{fu,}^n e_{fu}^n]$ and the lower membership function is represented by three points $[s_{fl}^n, t_{fl,}^n e_{fl}^n]$. Since we are using the triangular membership functions shown in Figure 11 then $t_{fu}^n = t_{fl}^n$ holds for all $\tilde{A}_f^n$, where $t_{fu}^n = t_{fl}^n$could be computed as follows:

$$t_{fu}^n = t_{fl}^n = step_n \times f + \min(x_n) \qquad (18)$$

To calculate s_{fu}^n and s_{fl}^n we use the following equations:

$$s_{fu}^n = start_f^n - dfou + \min(x_n) \qquad (19)$$
$$s_{fl}^n = start_f^n + dfou + \min(x_n) \qquad (20)$$

To calculate $e_{(f-1)u}^n$ and $e_{(f-1)l}^n$ we use the following equations:

$$e_{(f-1)l}^n = s_{fu}^n \qquad 2 < f \le F \qquad (21)$$
$$e_{(f-1)u}^n = s_{fl}^n \qquad 2 < f \le F \qquad (22)$$

In case $f = 1$, $t_{fu}^n = t_{fl}^n = \min(x_n)$ and in case $f = F$, $t_{fu}^n = t_{fl}^n = \max(x_n)$.

Case 2 ($F = 2$): In case F=2, all the equations above will apply. However startn and dfou calculations will change as noted below:

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

$$start^n = \frac{\max(x_n)-\min(x_n)}{2} \quad (23)$$

$$dfou = \frac{\max(x_n)-\min(x_n)}{4} \quad (24)$$

Figure 12 shows an example of one of the inputs' membership functions in case F=2.

(24)

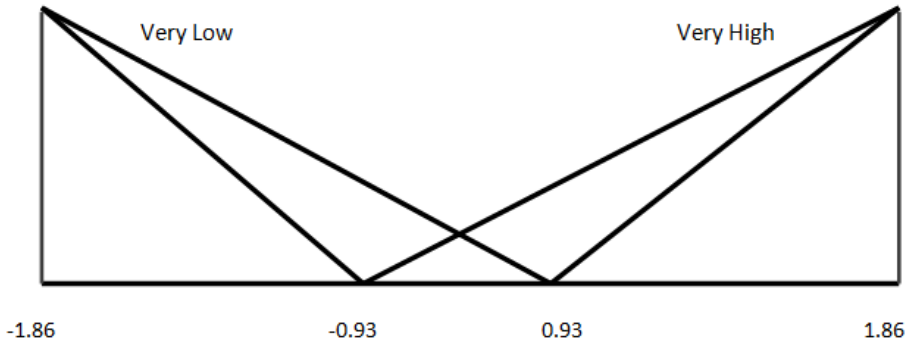


Fig.12: Example of one of the inputs' membership functions in case F= 2

Generating the Classifier fuzzy rules from training data

The set of interval type-2 Membership Functions generated from the previous step are combined with the accumulated user input/output data to extract the rules defining the classifier's behaviors. The rule extraction method is a one pass technique for extracting fuzzy rules from the data.

The type-2 hierarchical fuzzy logic classifier extracts single-input–single-output rules which describe the relationship between y (where y represents a class) and $x = (x_1, \dots, x_n)$ where each input in the data is taken and its rules get extracted with the single output so it can take the following form:

$$\text{IF } x_n \text{is } \hat{A}_n^i \text{ THEN y is O} \quad (25)$$

$i = 1,2, \dots, R$, where R is the number of rules and i is the index of the rule.$n = 1,2, \dots, N$, where N is the number of inputs and n is the index of the input. Finally there is O which represents the numerical value of the output (class) and in our case it can be -1 for (class 1) or +1 for (class 2) of the P300 BCI application.

After extracting the rules from the data, we remove the duplicated rules and by duplicated we mean the rules that share the same antecedent and we don't care about which consequent we choose as we will be optimizing the rules' consequents in the next steps using GA.

In this step, the extracted rule base is passed to a GA where the solution length is equal to the size of the extracted rules and each gene in the solution represents the consequent in the rule and the gene's index is equal to the rule's index as shown in Figure 13 assuming we have 10 rules only.

The initial solution is generated randomly where each gene can be one of these three values (-1, 0, 1) where -1 means that the output of this rule is class 1, 0 means that the classifier don't care about this rule and we won't take it into consideration and 1 means that the output of this rule is class 2.

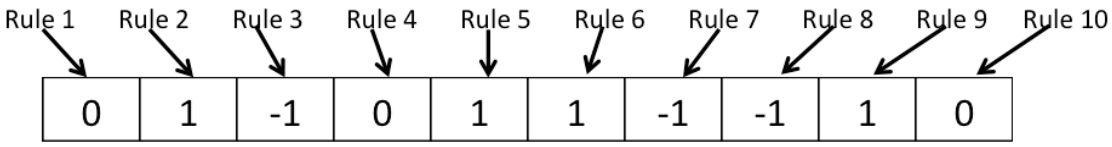


Fig.13: Example of the employed GA's chromosome

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

In this step, we use training data to tune the extracted rules and use the following fitness value:

$$\text{Fitness value} = \frac{\text{Correctly classfied class 1}+\text{Correctely classfied class 2}}{2} \quad (26)$$

In order to classify an input value, we use the following steps:

We get all the fired rules as follows:

We compute the upper and lower membership values $\overline{\mu}_{\tilde{A}_s^q}(x_s^{(t)})$ and $\underline{\mu}_{\tilde{A}_s^q}(x_s^{(t)})$ for each fuzzy set q=1,..Vi, and for each input variable s=1,…,N. Find q*∈{1,…Vi} such that :

$$\mu_{\tilde{A}_s^{q^*}}{}^{cg}(x_s^{(t)}) \geq \mu_{\tilde{A}_s^q}{}^{cg}(x_s^{(t)}) \text{ for all q=1,..Vi} \quad (27)$$

Where $\mu_{\tilde{A}_s^q}{}^{cg}(x_s^{(t)})$ is the center of gravity of the interval membership of $\tilde{A}_s^q$ at $x_s^{(t)}$ as follows (Hagras et al. 2007):

$$\mu_{\tilde{A}_s^q}{}^{cg}(x_s^{(t)}) = \frac{1}{2}\left[\overline{\mu}_{\tilde{A}_s^q}(x_s^{(t)}) + \underline{\mu}_{\tilde{A}_s^q}(x_s^{(t)})\right] \quad (28)$$

The firing strength of the rule $fl^{(t)}$ could be written as follows:

$$fl^{(t)} = \prod_{s=1}^{n} \mu_{\tilde{A}_s^q}{}^{cg}(x_s(t)) \quad (29)$$

We calculate the sum of the average firing strength of all the rules that have class 1 as a consequent (Let's call it α).

We calculate the sum of the average firing strength of all the rules that have class 2 as a consequent (Let's call it β).

If α is greater than β then the predicted class is class 1 otherwise the predicted class is class 2.

After every GA generation, we use the validation data to validate our best solution (tuned by training data) and the best validated solution is stored. We validate the solution using the same steps that were mentioned before to classify an input dataset. After the GA finished all its generations, we use the best validated solution to classify the testing data using a confusion matrix.

III. EXPERIMENTAL RESULTS

3.1 Results of the denoising method.

In this work, the data of (Hoffmann et al., 2008) was used to test the developed denoising method and compared it to both the case of no denoising and the case of wavelet shrinkage denoising(DL, 1995), (Saavedra & Bougrain, 2010). We followed the exact same sequence of pre-processing and classification in this paper to allow the direct comparison between the two cases of pre-processing with and without the denoising step. The description of the data set is found in detail in (Hoffmann et al., 2008) but a summary will be provided here. The duration of one run was approximately one minute and the duration of one session including setup of electrodes and short breaks between runs was approximately 30 min. One session comprised on average 810 trials, and the whole data for one subject consisted on average of 3240 trials. The impact of different electrode configurations and machine learning algorithms on classification accuracy was tested in an offline procedure. For each subject four-fold cross-validation was used to estimate average classification accuracy. The pre-processing operations applied were: referencing, bandpass filtering with cut-off frequencies set to 1.0 Hz and 12.0 Hz, downsampling by a factor of 64, single trials were extraction, windsorizing and finally amplitude normalization. The number of electrodes was selected as 4, 8, 16 or 32 depending on the experiment with the same electrode configurations in (Hoffmann et al., 2008). Then, the feature vector construction was done whereby the samples from the selected electrodes were concatenated into feature vectors. The dimensionality of the feature vectors was Ne × Nt, where Ne denotes the number of electrodes (selected as 4, 8, 16, or 32) and Nt denotes the number of temporal samples in one trial (32 samples in our experiments). Classification of data was performed using Bayesian linear discriminant analysis (BLDA) and the software developed by (Hoffmann et al., 2008) was used to perform this step. Given that the original signal passed through the standard pre-processing chain including the bandpass filter, comparing the results of different methods to it includes bandpass filter based denoising in the comparison. For the wavelet denoising, standard wavelet shrinkage denoising was used using Matlab with the basic wavelet chosen as "Coiflet-3" as suggested by (Saavedra & Bougrain, 2010) for direct comparison noting that we were able to get similar results using other basic wavelet functions (e.g., Daubechies-8). The universal threshold was selected with no multiplicative threshold rescaling (Saavedra & Bougrain, 2010).

3.2 Results of the Unsupervised Processing Method

The experimental P300-based BCI data of Hoffmann et al. (Hoffmann et al., 2008) were also used to test the developed no-training unsupervised methods and compare it to their results that were obtained with 3 sessions of training of a Bayesian Linear Discriminant Analysis (BLDA) classifier. To make that comparison directly applicable, we followed the exact same sequence of pre-processing and classification in this paper. The description of the data set is found in detail in Hoffmann et al. 3 but a summary will be provided here. The duration of one run was approximately one minute and the duration of one session including setup of electrodes and short breaks between runs was approximately 30 min.

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

One session comprised on average 810 trials, and the whole data for one subject consisted on average of 3240 trials. The experimental paradigm consists of flashing one of six images in a random order after asking the subject to count how many times a particular image appears. So, the six stimulus images appear in 6 consecutive trials, usually termed a block. The P300 signal is triggered by the appearance by the image of interest only (i.e., the one currently being counted and not the other five images) and can be detected from EEG signals to indicate the subject selection. In the supervised BLDA method, four-fold cross-validation was used to estimate average classification accuracy for each subject. So, each result from this classifier needs 3 sessions for training to compute. On the other hand, the proposed techniques work directly on the data without any prior training. This is a major difference between the previous methods and this work.

The standard pre-processing operations were applied to the data including referencing, bandpass filtering with cut-off frequencies set to 1.0 Hz and 12.0 Hz, downsampling by a factor of 64, single trials extraction, windsorizing and finally amplitude normalization. Additionally, for the new approach, signal denoising based on spectral subtraction was employed to the raw data before the above pre-processing(Kadah, 2004). Other methods were used based on wavelet denoising and other types of filters can also be used for similar results (Mustafa, Abrahim, Yassine, Zayed, & Kadah, 2012), (Abrahim, Mustafa, Yassine, Zayed, & Kadah, 2012). The denoising block is placed before the standard pre-processing steps above as shown in Figure 6. The number of electrodes was selected as 4, 8, 16 or 32 depending on the experiment with the same electrode configurations in the data set used (Hoffmann et al., 2008). Then, the samples from the selected electrodes were concatenated into feature vectors to be used for classification using either supervised BLDA (Hoffmann et al., 2008) or based on the new approach in this work. The dimensionality of the feature vectors was Ne × Nt, where Ne denotes the number of electrodes (selected as 4, 8, 16, or 32) and Nt denotes the number of temporal samples in one trial (32 samples in our experiments). The results of the different methods proposed are compared to each other and to supervised BLDA classification. The performance is measured using the block accuracy measure which is most relevant comparison criterion in this application. The block accuracy considers the data as blocks of 6 trials where only one of them should be selected with P300 signal showing while the others are not. If the classification results indicate anything other than only one activation at the correct image, it considers the whole block as incorrect. The results using different numbers of blocks were achieved by summing the signals from the selected number of blocks and using the sum as the new signal for classification using the proposed techniques. For the new approach, the block accuracy estimation experiments were repeated 24 times for independent sets of blocks containing trials from the same session and from different sessions for a given subject to avoid any bias and obtain accurate final results. The results are computed as block accuracy results for each subject, and average block accuracy results for all subjects. Also, relative block accuracy results were obtained by dividing the block accuracy results of the proposed methods by the block accuracy of the reference supervised BLDA method to allow better assessment of the performance.

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

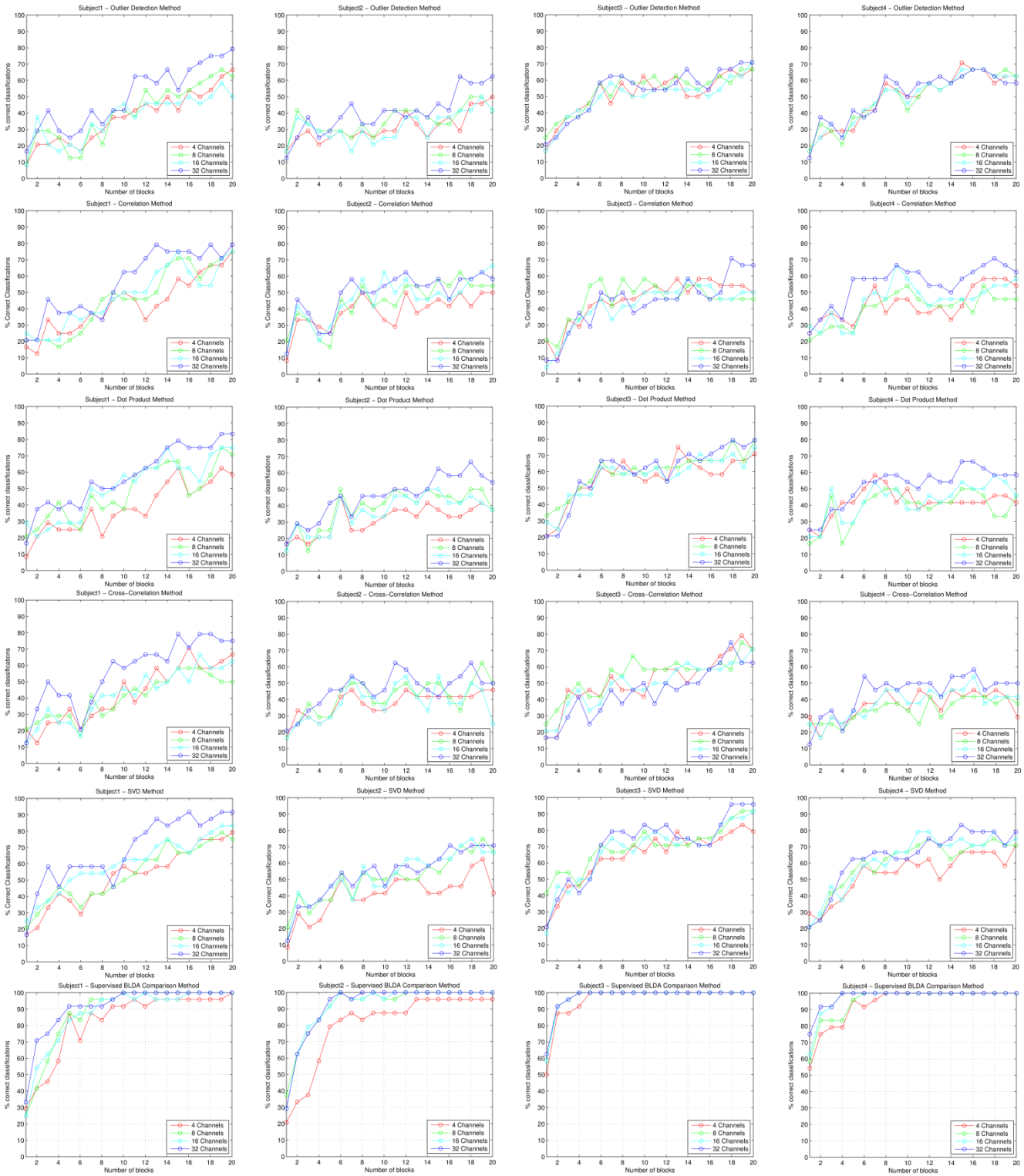
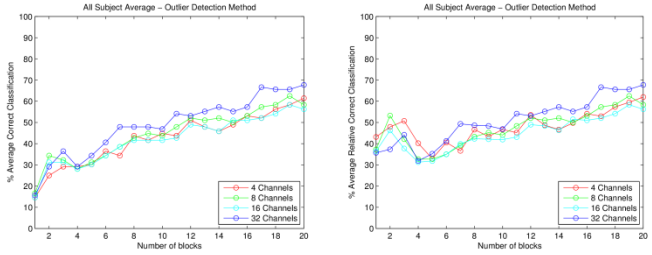


Fig.14:Block accuracy results for sample cases using all methods and electrode configurations.

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

Fig. 15: Relative block accuracy results for sample cases using all methods and electrode configurations

The block accuracy results of using the new approach on 4 sample subjects are shown in Figure 14. The figure presents the results using (a) outlier detection method, (b) correlation method, (c) dot product method, (d) cross correlation method, (e) singular value decomposition method, and (f) supervised classification using BLDA for direct comparison, each on a separate row.



https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

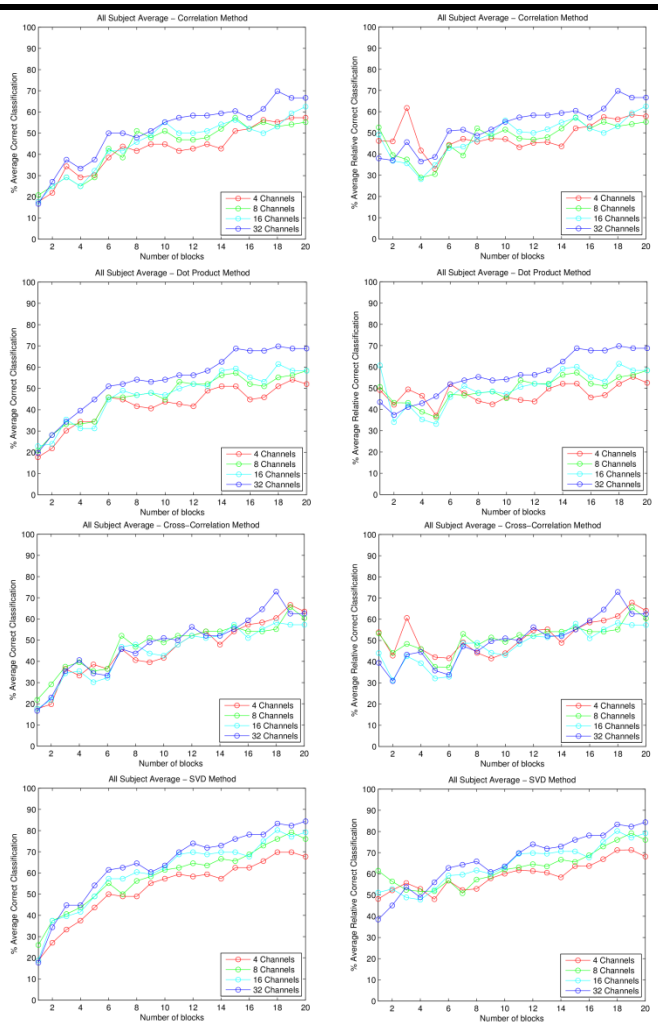


Fig.16: Average block accuracy and relative block accuracy results over all cases using all methods and electrode configurations

The results also show the cases of using 4, 8, 16 and 32 channel data on the same graph for each case/method. The relative block accuracy results are shown in Figure 15 with the same order of the methods for better interpretation. In Figure 16, the average block accuracies and relative block accuracies for all subjects are presented for each method to see an overall picture of the performance.

Table.1: Performance of different methods in terms of their low and high block classification accuracies computed over all subjects as compared to the results from supervised comparison method that requires 3-session training at the bottom row.

Block Accuracy Limits	**4-Channels**		**8-Channels**		**16-Channels**		**32-Channels**	
	Low	High	Low	High	Low	High	Low	High
Outlier Detection	14.6%	61.5%	16.7%	62.5%	14.6%	58.3%	15.6%	67.7%
Correlation	17.7%	57.3%	20.8%	57.3%	17.7%	62.5%	16.7%	69.8%
Dot Product	17.7%	54.2%	20.8%	58.3%	22.9%	61.5%	19.8%	69.8%
Cross-Correlation	17.7%	66.7%	21.9%	65.6%	17.7%	58.3%	16.7%	72.9%
SVD	18.8%	69.8%	26.0%	79.2%	18.8%	80.2%	17.7%	84.3%
Comparison Method	38.5%	98.9%	44.8%	100%	43.8%	100%	50.0%	100%

In Table 1, the low and high limits for the block accuracies for each method are given whereas those for relative block accuracies are given in Table 2. In the following, the analysis of the results according to different parameters is presented.

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

Table.2: Performance of different methods in terms of their relative low and high block classification accuracies computed over all subjects with reference to the results from supervised BLDA comparison method that requires 3-session training.

Block Accuracy Limits	4 Channels		8 Channels		16 Channels		32 Channels	
	Low	High	Low	High	Low	High	Low	High
Outlier Detection	32.9%	62.0%	32.7%	62.5%	31.3%	58.3%	31.9%	67.7%
Correlation	33.1%	61.8%	28.9%	57.3%	28.2%	62.5%	36.5%	69.8%
Dot Product	37.1%	55.3%	36.3%	58.3%	33.2%	61.5%	37.3%	69.8%
Cross-Correlation	41.5%	67.8%	37.3%	65.6%	31.2%	58.3%	30.9%	72.9%
SVD	**48.1%**	**71.3%**	**50.9%**	**79.2%**	**47.8%**	**80.2%**	**38.5%**	**84.4%**

- Effect of Method: The results from the sample individual cases show that the proposed method based on SVD provided the best performance reaching 95% block accuracy in some cases. The other 4 methods were comparable in block accuracy performance reaching accuracies above 80% in all cases. Examining the relative block accuracy curves, it is clear that all method range from 30% of the performance of the supervised BLDA method for low block averaging to above 90% in some cases with high block averaging and particularly for SVD. From the average performance curves and Tables 1 and2, the performance of the SVD method was clearly dominant on the high end of the range of block and relative block accuracies with average performance reaching 84.4%. The other proposed methods provide block and relative block accuracies around 70%. Their performances vary but they are within a close proximity of each other.
- Effect of number of channels: The difference is clear between the cases of 4 and 32 channels due to the implicit spatial averaging that occurs in using the higher number of channels. However, the difference in performance between the cases of 8 and 16 channels was not much and they both present accuracies in the middle between those of 4 and 32 channel data.
- Effect of number of blocks: There is a clear linear relationship between the block and relative block accuracies and the number of blocks used that is evident in all average curves. This is expected since the higher number of blocks allows more temporal averaging that improve the signal-to-noise ratio of the trials enhancing their separation procedures.

Variability among subjects: Some variations among subjects were observed where the results from Subject 2 for example were significantly lower than those from other subjects

3.3 Results of the Type-2 Fuzzy classifier

In this work, we will employ two data sets. The First data set is a standard data set obtained from (Hoffmann et al., 2008) where the data is a P300 application including a screen on which six images were displayed. The images were flashed in random sequences, one image at a time, the Inter-Stimulus Interval (ISI) was 400 ms.

The other data set was a real-world data set obtained from real subjects at the BCI lab in King Abdulaziz University (KAU). The data was again for a P300 application involving a screen on which six characters were displayed. The letters were flashed in random sequences, one raw\column at a time, ISI was 300 ms. Table 3 illustrates the differences between Hoffmann and KAU datasets. Figure 17 shows the KAU BCI lab setup and the subjects involved in the real world experiments.

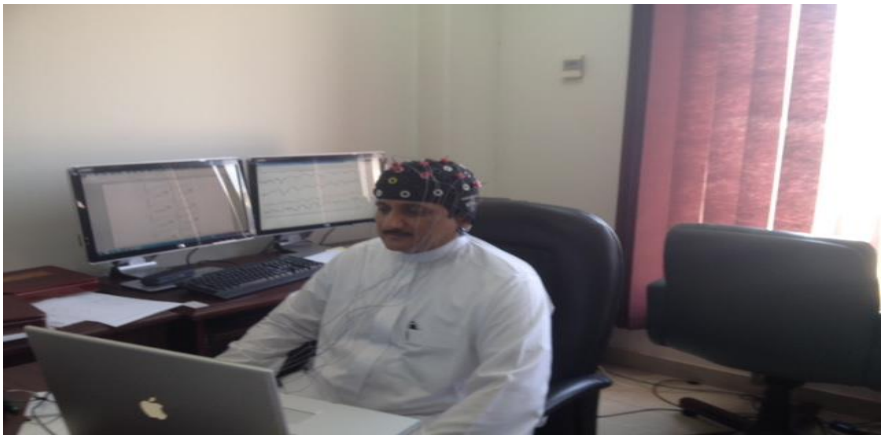


Fig.17: The KAU BCI lab setup and one of the subjects involved in the real world experiments.

Experimental Setup for Real-World Experiments

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

For the real world experiments, the users were facing a screen on which 6 Arabic characters were displayed. The characters were arranged in a 3 by 2 matrix, as shown in Figure 18.

Fig.18: The user display for our experiment

The user's task was to focus attention (i.e., count silently) on one character at a time in a word that was prescribed by the investigator. Table 4 illustrates the prescribed words. All rows and columns of this matrix were flashed in random sequences, one row/column at a time. Two out of 5 flashes of rows or columns contained the desired character (i.e., one particular row and one particular column).

For each character the matrix was displayed for a 2.5s period, and during this time the matrix was blank. Subsequently, each row and column in the matrix was randomly intensified for 100ms (i.e., resulting in 5 different stimuli, 3 rows and 2 columns). After intensification of a row/column, the matrix was blank for 200ms. Row/column intensifications were block randomized in blocks of 5. The sets of 5 intensifications were repeated 20 times for each character (i.e., any specific row/column was intensified 20 times and thus there were 100 total intensifications for each character). Each character was followed by a 2.5s period, and during this time the matrix was blank. During this period, the subject was instructed to spell the word أبتثجح given in Table 4. The subject will repeat the word four sessions. This period informed the user that this character was completed and the user needed to focus on the next character in the word that was displayed on the top of the screen (the current character was shown in parentheses).

The experiments were designed and recorded with BCI2000. The data for this experiment were collected from four normal males (26± 4.5 years). The EEG was recorded at 256 Hz sampling rate, with band pass filter from 0.1-60 Hz, and the notch filter was set on at 60Hz. The EEG was recorded using eight electrodes placed at the standard positions of the 10-20 international system. As shown in Figure 19, the selected electrodes were F3, F4, C3, Cz, C4, P3, Pz and P4 with AFz as ground and right ear lobe as reference.

As shown in Figure 20, the recording system consists of the following components: g.tecEEGcap, 8 Ag/AgCl electrodes, g.tecGAMMAbox, g.tecUSBamp and BCI2000.

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

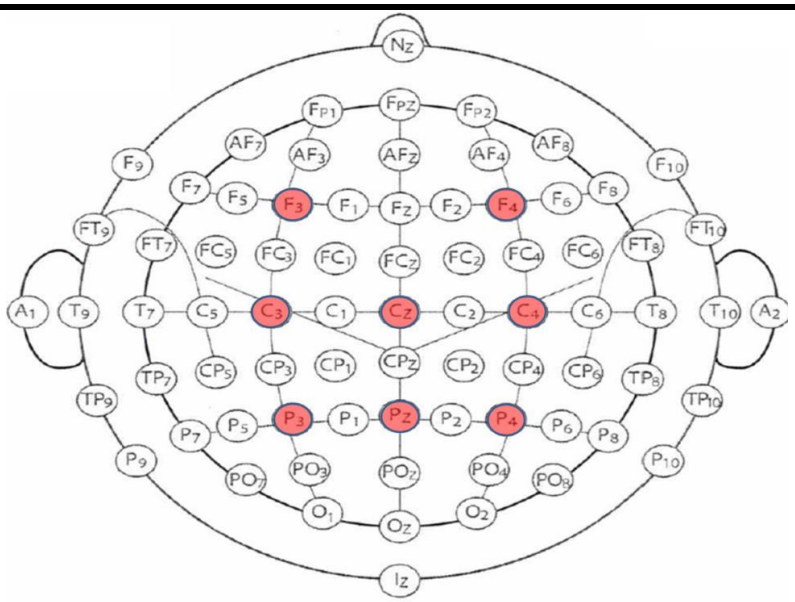


Fig.19: Selected electrodes for our experiment

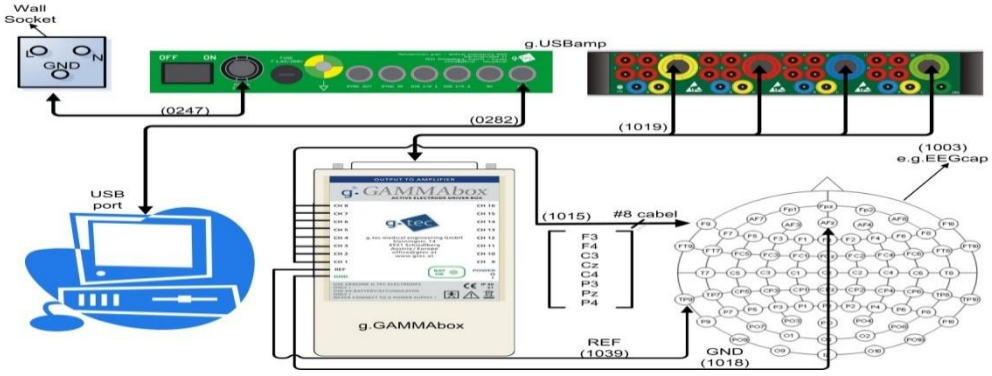


Fig.20: Simple 8 channels hardware diagram .

Data Pre-processing for KAU data 1

Before learning a classification function and before validation, several pre-processing operations were applied to the data. The main objective of this phase is to enhance the signal to noise ratio SNR. The pre-processing operations were applied in the order stated below.

- Referencing: During this phase twelve different re-reference techniques were applied where their results were compared with each other. The results showed that Common Average Reference (CAR) is best suited to be the reference technique. The twelve different re-reference techniques are listed below:
  - Common Reference: No re-montaging is done
  - Common average reference: The mean of all the electrodes is removed for all the electrodes .
  - Laplacian (4 adjacent): The weighted mean (depends on the distance) of the 4 adjacent electrodes is removed from the central electrode.
  - Surface Laplacian (8 adjacent): The weighted mean (depends on the distance) of the 8 surrounding electrodes is removed from the central electrode.
  - Bipolar (front to back): The difference of an electrode with the one behind it .
  - Bipolar (front to back skip 1): The difference of 2 electrodes that lies in front and also behind that electrode.
  - Bipolar (Symmetrical): The difference of 2 electrodes that is symmetrical to one another.
  - Bipolar (left to right): The difference of an electrode with the one right to it.
  - Bipolar (right to left): The difference of an electrode with the one left to it.
  - Using T7,T8 channels : The mean of T7,T8 channels is removed for all the electrodes .
  - Common average reference without mastoid channels: The mean of all the electrodes without mastoid channels is removed for all the electrodes .
  - Reference estimation:
- Filtering: A 6th order forward-backward Butterworth bandpass filter was used to filter the data. Cutoff frequencies were set to 1.0 Hz and 12.0 Hz. The MATLAB function butter was used to compute the

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

filter coefficients and the function filtfilt was used for filtering.

- Downsampling: The EEG was downsampled from 256 Hz to 32 Hz by selecting each 8th sample from the bandpass-filtered data.
- Single Trial Extraction: Single trials of duration 1000 ms were extracted from the data. Single trials started at stimulus onset, i.e. at the beginning of the flash of raw\column, and ended 1000 ms after stimulus onset.
- Winsorising: Winsorising or Winsorization is the transformation of statistics by limiting extreme values in the statistical data to reduce the effect of possibly spurious outliers. It is named after Charles P. Winsor (1895–1951). The effect is the same as clipping in signal processing. Eye blinks, eye movement, muscle activity, or subject movement can cause large amplitude outliers in the EEG. To reduce the effects of such outliers, the data from each electrode were clipped. For the samples from each electrode the 10th percentile and the 90th percentile were computed. Amplitude values lying below the 10th percentile or above the 90th percentile were then replaced by the 10th percentile or the 90th percentile, respectively.
- Normalization: The samples from each electrode were scaled to the interval [−1, 1]. The normalization was done using z-score method.
- Feature Vector Construction: The samples from the electrodes were concatenated into feature vectors. The dimensionality of the feature vectors was Ne × Ns × Nt, where Ne denotes the number of electrodes, Ns denotes the number of temporal samples in one trial, and Nt denotes the number of trials. Due to the trial duration of 1000 ms and the downsampling to 32 Hz, Ns always equaled 32. Depending on the electrode configuration, Ne equaled 8. Nt is varying according the number of trials in each session. After that the feature vectors is converted from three dimensions to two dimensions by concatenating the electrodes to each other. The dimensionality of the new feature vectors was Nes × Nt, where Nes equaled (8 × 32 = 256) and Nt is as is with no change.

Performance Evaluation Methodology for the Hofmann and KAU data

We have done all the training based on 70 % of (Hoffmann et al., 2008) data and we have used 30 % of the Hofmann data and the KAU as testing data on the generated classifiers. The training data was obtained from each of the 8 subjects in Hofmann data where 70 % of each individual data was used as training/validation and then the data was accumulated across the 8 subjects to form the training data. The remaining 30 % data for each subject of the data for was taken from each subject and accumulated across all subjects to result in the Hofmann testing data.

Table.3: The results on the Hofmann et al standard data set.

Number of Sensor Electrodes	BLDA			RFLDA			Type-2 Fuzzy Classifier		
	Positive	Negative	Average	Positive	Negative	Average	Positive	Negative	Average
4	58.64	55.30	56.97	44.81	65.87	55.34	61.12	56.83	58.98
8	59.38	56.12	57.75	48.04	68.45	58.24	54.31	63.42	58.87
32	50.37	62.47	56.42	33.28	78.74	56.013	58.91	58.94	58.93

Table 3 shows the achieved results on the standard Hofmann et al. standard data set over the testing data when using 4 sensors electrodes, 8 sensors electrodes and 32 electrodes respectively. It is shown the proposed type-2 fuzzy logic based classifiers results in better average accuracy of prediction when compared to the BLDA and RFLDA classifiers which are widely used in the BCI literature. The other advantage of the proposed type-2 fuzzy logic based classifier is that it produced a very small number of rules where the length of each rule is only one antecedent. Hence, this enables compact linguistic model which is easy to understand and analyses by the normal clinician. On the other hand, the BLDA and RFLDA classifiers are black box models which could not be understood and analyzed by the normal clinicians.

For example in case the 4 sensors electrodes, the generated rule base was composed of 38 rules (i.e. a rule for each input) of which 19 rules have an output class +1 and 19 rules having output class -1 as shown in Table 3, the generated rules are very simple with only one antecedent and there is a small number of rules. Hence, when compared to the BLDA and RFLDA the proposed type-2 fuzzy logic based classifier gave a better average prediction accuracy over the testing data of the Hoffmann data set. On the other hand, the proposed type-2 fuzzy logic classifier has resulted in an easy to understand and analyses linguistic model which explains the P300 phenomena over various subjects. The generated model could be easily understood and analysed by a normal clinician. Hence, the proposed type-2 fuzzy logic classifier can provide more understanding about the underlying processes and phenomena happening on the brain in a simplified way where for example rule 1 is

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

saying that the output will be Class +1 IF the signal from the Third Sensor at the 6th time instance is Low.

Performance Methodology for the Real-World KAU data

In order, to evaluate the performance of the BLDA, RFLDA and the suggested type-2 fuzzy logic classifier, we have tested the generated classifiers from the Hofmann data sets with 8 electrodes sensors on the unseen KAU data which was collected on different subjects and under different labs and sensor conditions.

As shown in Table 3, the proposed type-2 fuzzy logic based classifiers handled the uncertainties between the Hoffmann data set and the KAU data set which involved other subjects and other lab conditions and equipment where the type-2 fuzzy logic classifier has given a similar accuracy to the type-2 classifier over the Hoffmann data set. The type-2 fuzzy based classifier has given much better accuracy for the positive class, negative class and average accuracy when compared to the BLDA and RFLDA classifiers. It should be noted that the accuracies for the BLDA and RFLDA classifiers have degraded significantly from the Hofmann subjects to the KAU subjects as they are not capable to handle the uncertainties and they were trained to work for specific subjects and under certain lab conditions.

IV. DISCUSSION

It can be observed that the block accuracy results for 4-channel data (plotted in red) show a significant improvement from the original data in both spectral subtraction and wavelet shrinkage methods with low number of blocks. This is also reflected as higher bitrates in the same range. Even though the effect of denoising in general is more apparent in experiments with lower number of channels and low number of blocks, there is still evident improvement in experiments with high number of channels where 100% accuracy is reached earlier as evident in all cases. This is important to indicate that the inherent spatial compounding from the many electrodes can still take advantage of temporal denoising methods and that a combination of the two yields the best results.

By inspecting the results further, we observe that the spectral subtraction method offers better results than wavelet shrinkage based denoising in most experiments with the exception of a few cases such as in the 4-channel data of Subject 2 where the 100% accuracy is maintained once reached in wavelet denoising while it does not with spectral subtraction. Nevertheless, in all other cases the spectral subtraction results are superior as evident in the achieved block accuracy and bit rate for any given experiment. As a general observation, the results of spectral subtraction and wavelet denoising methods show a clear advantage over the results with only bandpass filtering in the original signal. Since such denoising step can be inserted within the conventional pre-processing of BCI data, this study shows clear evidence that these more sophisticated denoising methods should be integrated as a standard step in the pre-processing chain to improve the SNR of the collected signals.

Assuming a data set of M channels with N points each, the computational complexity of spectral subtraction is O(M N log2 N). On the other hand, The computational complexity of wavelet shrinkage method varies with different implementation with a minimum complexity of O(M N2), which is significantly higher. For example, for N=100000 points and same number of channels, the wavelet shrinkage method will require N/log2(N) times the computations of spectral subtraction, which is more than 3 orders of magnitude higher. Therefore, the computational complexity of spectral subtraction is more efficient for applications requiring embedded implementations or fpr real-time processing.

The model used in data processing amounts to subtracting the noise component uniformly across all frequencies. This is different from conventional frequency selective filters that are equivalent to a convolution in the time domain that causes the noise components in different time points to be correlated in the output signal. Hence, a theoretical advantage of this method is its preservation of the independence of random components within the time points processed. Hence, it is well-suited for use with standard statistical analysis methods that require statistical independence of samples. An example of such methods is when improving statistical estimation by using data from multiple blocks where the presence of correlated rather than independent noise across blocks degrades the achievable improvement. Given that the wavelet shrinkage based methods involve frequency selective filters to compute its coefficients, the same advantage cannot be claimed for that method. This explains the overwhelmingly better performance of the spectral subtraction method than the wavelet shrinkage based method when the number of blocks is higher.

From a global overview of results, one can observe that the new methods with no training requirement were able to achieve relative block accuracies of above 70% of the performance of the supervised BLDA method that require prior lengthy training with 3 full sessions. This is particularly important for such applications as P300-based BCI where the disabled person chooses one out of several images to indicate the need for a particular action. Such selection is usually done infrequently and with time separation that would require training to be repeated every time one selection has to be made, which would make this cumbersome for practical use. This demonstrates potential for the new approach that works adaptively without any

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

prior knowledge or assistance from caregivers or family members and without the training overhead required in supervised methods.

Given that the correct communication between the brain and the computer must include no ambiguity, a measure that considers the correct answer at the level of a whole block (i.e., one of six images) rather than an individual image on/off measure must be used. For example, if within a particular block 2 images out of 6 are classified as "selected" with only one of them a true selection, the usual accuracy would give a success rate of 5 out of 6, which is 83.3%. This is clearly incorrect because the message received was ambiguous. On the other hand, the block accuracy considers this whole block as incorrectly classified and would give a success rate of 0%, which is a realistic assessment of the utility of the received information. Other measures were used in other studies as well such as the bit rate. Here, given that we are comparing methods with no training to others with long training, the commonly used definition of bit rate is clearly flawed because it fails to account for the time needed for the required training and the fading of such training with time. Therefore, it was not possible to utilize this measure in this work.

The results for a particular number of blocks were achieved by summing the signals from the selected number of blocks and using the sum as the new signal for classification using the proposed techniques. Another approach that can be used is to calculate the proposed classification metrics from each block and then sum up all metrics from the desired number of blocks then make the classification decision based on this sum. The approach we used gave a better performance and hence was preferred over this alternative. The analysis of the problem shows that this is due to the nonlinearity of the computed measures that makes the average of the individual block measures completely different from the measure of average of blocks.

The applications of the new approach include developing plug-and-play P300 based BCI devices that require no training and work straight out of the box. Even though the block accuracy of such devices will be lower than the conventional methods with prior training, its adaptive nature and availability for immediate use without calibration boosts the robustness of their performance and practical utility.

V. CONCLUSION

In this work, a new denoising method for P300-based brain-computer interface data that allows better performance to be obtained with lower number of channels and blocks was developed. The new method was verified using experimental data and promising improved results were obtained. The new method was favorably compared to bandpass filtering and wavelet shrinkage based denoising as the present relevant and widely used method for denoising. Performance in different experiments using classification block accuracy as well as bit rate show significant improvement with a clear advantage in computational complexity. The results highlight the potential for including the new method as a standard pre-processing block for BCI data.

The results of a new approach for processing P300-based brain-computer interface data that allows classification of trials within a block without prior training are presented. The new method was verified using experimental data and compared to the results obtained with conventional processing with lengthy training. Promising results were obtained suggesting potential for the new approach in making the P300 based BCI technology easier to implement as plug-and-play device with no prior calibration required and capable of adaptively follow any changes in the subject's condition.

We presented a new system which can be regarded as a step towards understanding in easy linguistic formats, the complex phenomena occurring in the brain. The proposed system compromises a new feature selection mechanism which based on an ensemble neural network feature elimination system. The proposed feature selection system is capable of finding the most important time instances affecting each sensor in a P300 BCI application.

We have also presented a type-2 fuzzy logic based classifier which is able to handle the various uncertainties associated with the P300 BCI phenomena to produce better prediction accuracies than other competing classifiers such as BLDA or RFLDA. In addition, the generated type-2 classifier is learnt from data to produce very small number of rules with a rule length of only one antecedent to maximize the transparency and interpretability for the normal clinician.

We have presented various experiments which were performed on standard data sets and real-data sets obtained from the BCI lab in King Abdulaziz University. It was shown that the produced type-2 fuzzy logic based classifier learnt simple rules which are easy to understand explaining the events in question. In addition, the produced type-2 classifier was able to give a better average accuracy than BLDA or RFLDA on various human subjects on the standard data sets and on the real-world data sets. In addition, the proposed type-2 fuzzy classifier was able handle the uncertainties existing between the various human subjects and under the different lab and equipment conditions thus resulting in general models explaining the given phenomena rather than other classification methods (as in the case of BLDA or RFLDA) which are specific to certain human subjects and under certain lab conditions.

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

For our future work, we aim to explore general type-2 fuzzy logic classifiers in order to improve the accuracy of prediction of the type-2 fuzzy classifiers.

ACKNOWLEDGEMENT

We would like to thank our team for their efforts in the BCI project. This research was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Saudi Arabia, Jeddah, under grant, No. (16-15-1432 HiCi). The authors, therefore, acknowledge with thanks DSR technical and financial support.

REFERENCES

Abrahim, B. A., Mustafa, Z. A., Yassine, I. A., Zayed, N., & Kadah, Y. M. (2012). Hybrid total variation and wavelet thresholding speckle reduction for medical ultrasound imaging. *Journal of Medical Imaging and Health Informatics, 2*(2), 114-124.

Ahmadi, M., & Quiroga, R. Q. (2013). Automatic denoising of single-trial evoked potentials. *Neuroimage, 66*, 672-680.

Akhtar, M. T., Mitsuhashi, W., & James, C. J. (2012). Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data. *Signal Processing, 92*(2), 401-416.

Alhaddad, M. J., & , K. M., Malibary H. (2011). Unsupervised Adaptive P300 BCI in the framework of chaotic theory and stochastic theory, First Technical Report BCI (Vol. 1): King Abdulaziz University, Collage of Computing and information Technology.

Alhaddad, M. J., Kamel, M., & Al-Otaibi, N. (2013). Comparative study of data fusion algorithms in P300 Based BCI. *Advances in Computer Science: an International Journal, 2*(4), 47-54.

Alhaddad, M. J., Kamel, M., & Al-Otaibi, N. (2014). Toward An Adaptive P300 Based Brain Computer. *World of Computer Science & Information Technology Journal, 4*(2).

Alhaddad, M. J., & Kamel M, M. H. (2012). Unsupervised Adaptive P300 BCI in the framework of chaotic theory and stochastic theory, Second Tech. Report. (Vol. 2): King Abdulaziz University, Collage of Computing and information Technology. .

Alhaddad, M. J., Kamel, M. I., & Bakheet, D. M. (2013). *Quran Player Based on Brain Computer Interface.* Paper presented at the Proceedings of the 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences.

Alhaddad, M. J., Kamel, M. I., & Bakheet, D. M. (2014). *P300-brain computer interface based on ordinal analysis of time series.* Paper presented at the 2014 International Conference on Computational Science and Computational Intelligence (CSCI).

Alhaddad, M. J., Kamel, M. I., & Kadah, Y. M. (2014). Investigation of New Unsupervised Processing Methods for P300-Based Brain-Computer Interface. *Journal of Medical Imaging and Health Informatics, 4*(3), 363-369.

Alhaddad, M. J., Kamel, M. I., Makary, M. M., Hargas, H., & Kadah, Y. M. (2014). Spectral subtraction denoising preprocessing block to improve P300-based brain-computer interfacing. *Biomedical engineering online, 13*(1), 36.

Alhaddad, M. J., Mohammed, A., Kamel, M., & Hagras, H. (2015). A genetic interval type-2 fuzzy logic-based approach for generating interpretable linguistic models for the brain P300 phenomena recorded via brain–computer interfaces. *Soft Computing, 19*(4), 1019-1035.

Candè, E. J., & Wakin, M. B. (2008). An introduction to compressive sampling. *Signal Processing Magazine, IEEE, 25*(2), 21-30.

Croux, C., Filzmoser, P., & Joossens, K. (2008). Classification efficiencies for robust linear discriminant analysis. *Statistica Sinica*, 581-599.

de Cheveigne, A., & Simon, J. Z. (2008). Denoising based on spatial filtering. *J Neurosci Methods, 171*(2), 331-339. doi: 10.1016/j.jneumeth.2008.03.015

DL, D. (1995). Denoising by soft-thresholding. *IEEE Trans Inf Theory, 42*, 613-627.

Effern, A., Lehnertz, K., Grunwald, T., Fernández, G., David, P., & Elger, C. E. (2000). Time adaptive denoising of single trial event- related potentials in the wavelet domain. *Psychophysiology, 37*(6), 859-865.

Estrada, E., Nazeran, H., Sierra, G., Ebrahimi, F., & Setarehdan, S. K. (2011). *Wavelet-based EEG denoising for automatic sleep stage classification.* Paper presented at the Electrical Communications and Computers (CONIELECOMP), 2011 21st International Conference on.

Gao, J., Sultan, H., Hu, J., & Tung, W.-W. (2010). Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: a comparison. *Signal Processing Letters, IEEE, 17*(3), 237-240.

Geetha, G., & Geethalakshmi, S. (2012). Artifact Removal from EEG using Spatially Constrained Independent Component Analysis and Wavelet

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

Denoising with Otsu's Thresholding Technique. *Procedia Engineering, 30*, 1064-1071.

Golub, G. H., & Van Loan, C. F. (2012). *Matrix computations* (Vol. 3): JHU Press.

Hammad, S., Corazzol, M., Kamavuako, E. N., & Jensen, W. (2012). *Wavelet denoising and ANN/SVM decoding of a self-paced forelimb movement based on multi-unit intra-cortical signals in rats.* Paper presented at the Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on.

Hammon, P. S., & De Sa, V. R. (2007). Preprocessing and meta-classification for brain-computer interfaces. *Biomedical Engineering, IEEE Transactions on, 54*(3), 518-525.

Herman, P., Prasad, G., & McGinnity, T. M. (2008). *Designing a robust type-2 fuzzy logic classifier for non-stationary systems with application in brain-computer interfacing.* Paper presented at the Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on.

Hoffmann, U., Vesin, J.-M., Ebrahimi, T., & Diserens, K. (2008). An efficient P300-based brain–computer interface for disabled subjects. *J Neurosci Methods, 167*(1), 115-125.

Hyviirinen, A., Karhunen, J., & Oja, E. (2001). Independent component analysis. *Wileyand Sons*.

intendiX. from http://www.intendix.com/

Kadah, Y. M. (2004). Adaptive denoising of event-related functional magnetic resonance imaging data using spectral subtraction. *Biomedical Engineering, IEEE Transactions on, 51*(11), 1944-1953.

Krauledat, M., Tangermann, M., Blankertz, B., & Müller, K.-R. (2008). Towards zero training for brain-computer interfacing. *PloS one, 3*(8), e2967.

Kübler, A., & Müller, K. (2007). An introduction to brain computer interfacing in Towards Brain-Computer Interfacing: MIT Press Cambridge.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering, 4*(2), R1.

Lotte, F., Lécuyer, A., & Arnaldi, B. (2009). FuRIA: an inverse solution based feature extraction algorithm using fuzzy set theory for brain–computer interfaces. *Signal Processing, IEEE Transactions on, 57*(8), 3253-3263.

Lotte, F., Lécuyer, A., & Lamarche, F. (2007). Studying the use of fuzzy inference systems for motor imagery classification. *IEEE transactions on neural systems and rehabilitation engineering, 15*(2), 322-324.

Mirghasemi, H., Shamsollahi, M., & Fazel-Rezai, R. (2006). *Assessment of preprocessing on classifiers used in the P300 speller paradigm.* Paper presented at the Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE.

Müller, K.-R., Anderson, C. W., & Birch, G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 11*(2), 165-169.

Mustafa, Z. A., Abrahim, B. A., Yassine, I. A., Zayed, N., & Kadah, Y. M. (2012). Wavelet domain bilateral filtering with subband mixing for magnetic resonance image enhancement. *Journal of Medical Imaging and Health Informatics, 2*(3), 230-237.

Palaniappan, R., Paramesran, R., Nishida, S., & Saiwaki, N. (2002). A new brain-computer interface design using fuzzy ARTMAP. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 10*(3), 140-148.

Pires, G., Nunes, U., & Castelo-Branco, M. (2011). Statistical spatial filtering for a P300-based BCI: tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis. *J Neurosci Methods, 195*(2), 270-281. doi: 10.1016/j.jneumeth.2010.11.016

Quiroga, R. Q., & Garcia, H. (2003). Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology, 114*(2), 376-390.

Ramirez, R. R., Kopell, B. H., Butson, C. R., Hiner, B. C., & Baillet, S. (2011). Spectral signal space projection algorithm for frequency domain MEG and EEG denoising, whitening, and source imaging. *Neuroimage, 56*(1), 78-92. doi: 10.1016/j.neuroimage.2011.02.002

Saavedra, C., & Bougrain, L. (2010). *Wavelet denoising for P300 single-trial detection.* Paper presented at the Proccedings of the 5th french conference on computational neuroscience-Neurocomp'10.

Sammaiah, A., Narsimha, B., Suresh, E., & Reddy, M. S. (2011). *On the performance of wavelet transform improving Eye blink detections for BCI.* Paper presented at the Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on.

Strang, G., & Press, W.-C. (2009). *Introduction to linear algebra* (Vol. 4): Wellesley-Cambridge Press Wellesley, MA.

Tu, Y., Huang, G., Hung, Y. S., Hu, L., Hu, Y., & Zhang, Z. (2013). *Single-trial detection of visual evoked*

https://dx.doi.org/10.22161/ijaers/3.12.34 *ISSN: 2349-6495(P) | 2456-1908(O)*

potentials by common spatial patterns and wavelet filtering for brain-computer interface. Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE.

Vázquez, R. R., Velez-Perez, H., Ranta, R., Dorr, V. L., Maquin, D., & Maillard, L. (2012). Blind source separation, wavelet denoising and discriminant analysis for EEG artefacts and noise cancelling. *Biomedical Signal Processing and Control, 7*(4), 389-400.

Vorobyov, S., & Cichocki, A. (2002). Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis. *Biol Cybern, 86*(4), 293-303. doi: 10.1007/s00422-001-0298-6

Windeatt, T., Duangsoithong, R., & Smith, R. (2011). Embedded feature ranking for ensemble MLP classifiers. *IEEE Transactions on Neural Networks, 22*(6), 988-994.